



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 113470

TO: Georgia L Helmer
Location: REM/2C15/2C18
Art Unit: 1638
Thursday, February 05, 2004

Case Serial Number: 09643755

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-B55
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner Helmer,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
CM-1, Rm. 6-A-06
605-1155

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 01:29:45 ; Search time 4588 Seconds
(without alignments)
10459.240 Million cell updates/sec

Title: US-09-643-755B-1
Perfect score: 1173
Sequence: 1 atgaacttcctaagctctt.....ggctagctaagaactctga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hngo_hum:*
40: em_hngo_mus:*
41: em_hngo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	100.0	1173	6	AX088019 Sequence
2	1173	100.0	3957	6	AX088021 Sequence
3	900.4	76.8	1305	4	BOVCHTMOB
4	897.2	76.5	1240	6	AR002347
5	897.2	76.5	1460	6	E00042
6	896.2	76.4	1291	6	A15836
7	895.6	76.4	1175	6	I04058
8	894.8	76.3	2733	6	AR073077
9	894	76.2	1275	4	BOVCHTMOA
10	894	76.2	1290	6	E00075
11	894	76.2	1291	6	A15633
12	893	76.1	1289	6	E00144
13	893	76.1	2726	6	I08097
14	892.4	76.1	1311	6	E00108
15	892.4	76.1	1460	6	E00295
16	890.8	75.9	1269	4	BOVCHTMOB
17	886.8	75.6	1098	6	E00079
18	886.8	75.6	1098	6	E02341
19	872.8	74.4	2982	6	I08098
20	870.8	74.2	1101	4	AR177290
21	839.6	71.6	1292	4	OAPCHY
22	776	66.2	1115	12	SYNPROCA
23	753.6	64.2	1094	4	BP19786
24	749	63.9	1172	4	CDR131677
25	732.4	62.4	1230	4	SSU14406
26	718	61.2	1262	9	AB038386
27	702	59.8	1095	4	AR421161
28	681.6	58.1	1080	4	AR421162
29	627.2	53.5	969	6	A02000
30	627.2	53.5	969	6	A06446
31	619.6	52.8	1146	12	SYNPROCA
32	605	51.6	1259	10	RN025168
33	569	48.5	957	4	AR421164
34	534	45.5	1140	6	AX19837
35	490.8	41.8	981	4	AR421163
36	382.4	32.6	637	6	E05472
37	381.8	32.5	1403	4	AB047246
38	378.6	32.3	1161	4	AB047245
39	373.8	31.9	858	4	AR421165
40	373.8	31.9	1351	9	AB038384
41	370.8	31.6	1381	4	AB047244
42	363.8	31.0	1362	4	AB047243
43	363.8	31.0	1363	4	PICPERA
44	363.4	31.0	1221	4	CDR131678
45	362.2	30.9	1276	5	CHKPSN

ALIGNMENTS

RESULT 1
AX088019 1173 bp DNA linear PAT 17-MAR-2001
LOCUS Sequence 1 from Patent WO0114571.
ACCESSION AX088019
VERSION AX088019.1 GI:13396947
KEYWORDS
SOURCE
ORGANISM
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.
TITLE Commercial production of chymosin in plants

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0114571-A 1 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES Location/Qualifiers
source 1..1173
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
1..1173
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC34680.1"
/db_xref="GI:13396948"
/translation="MNFLKSPFPYAFLCRGQYFVAVTHAAETRIPLYYGKSLRKALK
EHLLEDFLQKQOYGISSKYSGFEVASVPLTNVLDOSYFGKTYLGTPEQEFVLADT
GSSDFWVPSIYCKSNACKKHQRPDPKRSFQNLGKPLSHVGTGSMOGLIGDYTV
SNIVDIQOYVGLSTOBPGDVFYAEFGDILGMAYPSLASEYSIPVDNMNRLVAD
LFSVYDRNGOESMLTGLAIDPSYTGSLHWVPTVQOYQFTVDSVTISGVVACEG
GCOALIDTGSTKLVPSSDIINIQAIGATONOGEPDIDCDNLSTMPVVPFINCKM
YPLTBSATYSQDGFCSTSGFOSBNHOKMTLGDVFTREYISVFDNRANLVGLAKAI"

BASE COUNT 299 a 308 c 262 g 304 t
ORIGIN

Query Match 100.0%; Score 1173; DB 6; Length 1173;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTCTTAAGTCTTTCCCTTTCTGACGCTTTCCTTTGTTCCGTCATACTTCGTT 60
Db 1 ATGAACCTCTTAAGTCTTTCCCTTTCTGACGCTTTCCTTTGTTCCGTCATACTTCGTT 60

QY 61 GCTGTACTACAGCGCTGCTGATCACCGGATTCCTCTCTACAAAGTAGTCTCTCCGT 120
Db 61 GCTGTACTACAGCGCTGCTGATCACCGGATTCCTCTCTACAAAGTAGTCTCTCCGT 120

QY 121 AAGGCCCTGAGAGAACATGACTTCTTGAAGACTTCTTGGAGAAACAAGATAGCATC 180
Db 121 AAGGCCCTGAGAGAACATGACTTCTTGAAGACTTCTTGGAGAAACAAGATAGCATC 180

QY 181 AAGCAGCAAGTACTCCGGCTTCGGTGAAGTGTAGCGTGCACCTTACCACTACCTTGAT 240
Db 181 AAGCAGCAAGTACTCCGGCTTCGGTGAAGTGTAGCGTGCACCTTACCACTACCTTGAT 240

QY 241 AGTCATTAATCTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGAGTTCACCGTTCTTT 300
Db 241 AGTCATTAATCTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGAGTTCACCGTTCTTT 300

QY 301 GATACGTGTTCTCTGACTCTCGGTTCCCTCTATCTACTGCAAGAGCAATGCTGCAAG 360
Db 301 GATACGTGTTCTCTGACTCTCGGTTCCCTCTATCTACTGCAAGAGCAATGCTGCAAG 360

QY 361 AACCAACCAAGATTCGATCCGAGAAAGTGTCACTTCAGAACTTAGGCAAAACCTTTG 420
Db 361 AACCAACCAAGATTCGATCCGAGAAAGTGTCACTTCAGAACTTAGGCAAAACCTTTG 420

QY 421 TCTATACACTACGGGTACAGGTAGTACGAAGAACTTGAAGCTATGATACCGTCACTGTC 480
Db 421 TCTATACACTACGGGTACAGGTAGTACGAAGAACTTGAAGCTATGATACCGTCACTGTC 480

QY 481 TCCAACTTGTGACATTCGAAAGACAGAGTGAAGCTTACGACCCCAAGAAACAAGGTATGTC 540
Db 481 TCCAACTTGTGACATTCGAAAGACAGAGTGAAGCTTACGACCCCAAGAAACAAGGTATGTC 540

QY 541 TTCACCTATGACAGATTCGATGCGATCCTTGTATGAGCATACCATCGCTCGGCTGAGAG 600
Db 541 TTCACCTATGACAGATTCGATGCGATCCTTGTATGAGCATACCATCGCTCGGCTGAGAG 600

QY 601 TACTTCGATACCTGTGTTGACAAATGATGAAACCGACACTAGTAGCTCAAGACTTGTTC 660
Db 601 TACTTCGATACCTGTGTTGACAAATGATGAAACCGACACTAGTAGCTCAAGACTTGTTC 660

QY 661 TCGGTTTACATGACAGGAATGGCCAGAGAGATGCTCAAGCTTGGAGCTATTGATCCA 720
Db 661 TCGGTTTACATGACAGGAATGGCCAGAGAGATGCTCAAGCTTGGAGCTATTGATCCA 720

QY 721 TCTTACTACACAGGATCTCTTCACTGCGTTCCAGTCACTGTGACAGATGCTGGCAATTC 780
Db 721 TCTTACTACACAGGATCTCTTCACTGCGTTCCAGTCACTGTGACAGATGCTGGCAATTC 780

QY 781 ACTGTGACAGTGTACCAATCAGCGGTGTGTTGTTGACATGTTGAAGTGGATGTCAAGCT 840
Db 781 ACTGTGACAGTGTACCAATCAGCGGTGTGTTGTTGACATGTTGAAGTGGATGTCAAGCT 840

QY 841 ATCTTGATACCGGTAAGTCAAGTGTGTCGACCTTAGACGACCAATTCACATTCAG 900
Db 841 ATCTTGATACCGGTAAGTCAAGTGTGTCGACCTTAGACGACCAATTCACATTCAG 900

QY 901 CAAGCTATTGAGCCACACAGAAACAGTACGTTGAGTTGACATTAATTGCCAATCTT 960
Db 901 CAAGCTATTGAGCCACACAGAAACAGTACGTTGAGTTGACATTAATTGCCAATCTT 960

QY 961 AGCTATATGCTCAATGCTCTTTGATGATCAACGGCAAGATGATACCATGACCCCTCC 1020
Db 961 AGCTATATGCTCAATGCTCTTTGATGATCAACGGCAAGATGATACCATGACCCCTCC 1020

QY 1021 GCTTATACGACGACGATCAAGGTTCTGACACGATGATTCAGATGAGAACCATTC 1080
Db 1021 GCTTATACGACGACGATCAAGGTTCTGACACGATGATTCAGATGAGAACCATTC 1080

QY 1081 CAGAAATGATCTTGGGAATGTTCATTCTGATGATCAAGCGTCTTTGACAGGCTC 1140
Db 1081 CAGAAATGATCTTGGGAATGTTCATTCTGATGATCAAGCGTCTTTGACAGGCTC 1140

QY 1141 AACACCTGTTGGGCTAGCTAAAGCAATCTGA 1173
Db 1141 AACACCTGTTGGGCTAGCTAAAGCAATCTGA 1173

RESULT 2
AX088021
LOCUS AX088021 3957 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114571.
ACCESSION AX088021
VERSION AX088021.1 GI:13396949
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.

REFERENCE
1
AUTHORS van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.
TITLE Commercial production of chymosin in plants
JOURNAL Patent: WO 0114571-A 3 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES Location/Qualifiers
source 1..3957
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
1554..2726
/note="Figure 2"
/codon_start=1
/transl_table=11
/protein_id="CAC34680.1"
/db_xref="GI:13396950"
/translation="MNFLKSPFPYAFLCRGQYFVAVTHAAETRIPLYYGKSLRKALK
EHLLEDFLQKQOYGISSKYSGFEVASVPLTNVLDOSYFGKTYLGTPEQEFVLADT
GSSDFWVPSIYCKSNACKKHQRPDPKRSFQNLGKPLSHVGTGSMOGLIGDYTV
SNIVDIQOYVGLSTOBPGDVFYAEFGDILGMAYPSLASEYSIPVDNMNRLVAD
LFSVYDRNGOESMLTGLAIDPSYTGSLHWVPTVQOYQFTVDSVTISGVVACEG
GCOALIDTGSTKLVPSSDIINIQAIGATONOGEPDIDCDNLSTMPVVPFINCKM
YPLTBSATYSQDGFCSTSGFOSBNHOKMTLGDVFTREYISVFDNRANLVGLAKAI"

BASE COUNT 1263 a 790 c 609 g 1295 t
ORIGIN

Query Match 100.0%; Score 1173; DB 6; Length 3957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATTCCTTAAGTCTTCCCTTCTACGCTTCTTGTGTTGGTCAATACCTCTT 60
 Db 1554 ATGAATTCCTTAAGTCTTCCCTTCTACGCTTCTTGTGTTGGTCAATACCTCTT 1613
 QY 61 GCTGTACTCAGCGCTGTGAGATCCCGCATTCCTCTTCAAGAGTAAGTCTCTCGT 120
 Db 1614 GCTGTACTCAGCGCTGTGAGATCCCGCATTCCTCTTCAAGAGTAAGTCTCTCGT 1673
 QY 121 AAGGGCTGAAGAGCATGAGCTTGTAGAAGCTCTTGAGAAACAAGATGGCATC 180
 Db 1674 AAGGGCTGAAGAGCATGAGCTTGTAGAAGCTCTTGAGAAACAAGATGGCATC 1733
 QY 181 AGCAGCAAGTACTCGGCTGTGAGATGGTGTAGCGCCTTACCACTACCTTGAAT 240
 Db 1734 AGCAGCAAGTACTCGGCTGTGAGATGGTGTAGCGCCTTACCACTACCTTGAAT 1793
 QY 241 AGTCATATCTTGGGAGATCTACCTCGGAACCCGCTCAAGAGTTCACCGTCTCTT 300
 Db 1794 AGTCATATCTTGGGAGATCTACCTCGGAACCCGCTCAAGAGTTCACCGTCTCTT 1853
 QY 301 GATATGCTTCTCTGACTTCTGGGTTCCCTTATCTACTGCAAGCAATGCTTGAAG 360
 Db 1854 GATATGCTTCTCTGACTTCTGGGTTCCCTTATCTACTGCAAGCAATGCTTGAAG 1913
 QY 361 AACCAACCAAGATTCGATCCGGAAGTGTGACCTTCCAGACTTATGGCAACCCCTT 420
 Db 1914 AACCAACCAAGATTCGATCCGGAAGTGTGACCTTCCAGACTTATGGCAACCCCTT 1973
 QY 421 TCTATACACTACGATACAGTATGATGCAAGATCTTATAGCTATGATACCTGCTGTC 480
 Db 1974 TCTATACACTACGATACAGTATGATGCAAGATCTTATAGCTATGATACCTGCTGTC 2033
 QY 481 TCCAAATTTGTGACATTCACAGACAGTAGAGACTTATGACCCCAAGAACAGTATGTC 540
 Db 2034 TCCAAATTTGTGACATTCACAGACAGTAGAGACTTATGACCCCAAGAACAGTATGTC 2093
 QY 541 TTCACCTTATGAGATTCGATGGGATCCCTGTTATGGGATACCAATGCTGCGCTGAG 600
 Db 2094 TTCACCTTATGAGATTCGATGGGATCCCTGTTATGGGATACCAATGCTGCGCTGAG 2153
 QY 601 TACTGATATCCTGTGTGACCAATGATGAAACCAACCACTATGATCTCAAGACTTGTTC 660
 Db 2154 TACTGATATCCTGTGTGACCAATGATGAAACCAACCACTATGATCTCAAGACTTGTTC 2213
 QY 661 TCGGTTTACATGACAGAGATGGCCAGAGAGATGCTCAAGCTTGGAGCTATTTATCCA 720
 Db 2214 TCGGTTTACATGACAGAGATGGCCAGAGAGATGCTCAAGCTTGGAGCTATTTATCCA 2273
 QY 721 TCCCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGATGCTGGCAATTC 780
 Db 2274 TCCCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGATGCTGGCAATTC 2333
 QY 781 ACTGTGACAGAGTCAATCAGCAGCGGTGTGTGTGACATGAGAGTGAATGTCAGCT 840
 Db 2334 ACTGTGACAGAGTCAATCAGCAGCGGTGTGTGTGACATGAGAGTGAATGTCAGCT 2393
 QY 841 ATCTTGATACCGGATCGTCCAGCTGTGAGACCTTACGAGGACGACATTTCTCAATTCAG 900
 Db 2394 ATCTTGATACCGGATCGTCCAGCTGTGAGACCTTACGAGGACGACATTTCTCAATTCAG 2453
 QY 901 CAAAGCTATTTGAGCCACACAGAACAGTACGCTGATTTGACATGATTTGCCAACACTT 960
 Db 2454 CAAAGCTATTTGAGCCACACAGAACAGTACGCTGATTTGACATGATTTGCCAACACTT 2513
 QY 961 AGCTATCATCTCAAGTGTCTTTGATCAACGCAAGATGATCCCACTGACCCCTCTC 1020
 Db 2514 AGCTATCATCTCAAGTGTCTTTGATCAACGCAAGATGATCCCACTGACCCCTCTC 2573
 QY 1021 GCCTATACCAAGCAGAGATCAAGGGTTCTGACACAGTGAATTCAGAGTGAACAATTC 1080
 Db 2574 GCCTATACCAAGCAGAGATCAAGGGTTCTGACACAGTGAATTCAGAGTGAACAATTC 2633

QY 1081 CAGAAATGATCTTGGAGATGTGTTCATTTGTGATATCAACAGCTCTTTGACAGGCC 1140
 Db 2634 CAGAAATGATCTTGGAGATGTGTTCATTTGTGATATCAACAGCTCTTTGACAGGCC 2693
 QY 1141 AACCACTCTGTTGGGCTATGCTTAAGCAATCTGA 1173
 Db 2694 AACCACTCTGTTGGGCTATGCTTAAGCAATCTGA 2726
 RESULT 3
 BOVCHYMOSIN 1305 bp mRNA linear MAM 26-APR-1993
 LOCUS bovine chymosin b (rennin) mRNA.
 DEFINITION J00003
 ACCESSION J00003.1 GI:162859
 VERSION chymosin; chymosin B; protease; rennin.
 KEYWORDS SOURCE
 ORGANISM Bos taurus (cow)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 1 (bases 1 to 1305)
 Harris,T.J., Lowe,P.A., Lyons,A., Thomas,P.G., Eaton,M.A., Millican,T.A., Patel,T.P., Bose,C.C., Carey,N.H. and Doel,M.T.
 Molecular cloning and nucleotide sequence of cDNA coding for calf preprochymosin
 Nucleic Acids Res. 10 (7), 2177-2187 (1982)
 JOURNAL MEDLINE PUBMED
 82221400 6283469
 COMMENT Original source text: bovine (calf) cdna of fourth stomach mucosa mra.
 Chymosin is the major proteolytic enzyme in the fourth stomach of the unweaned calf. Two chromatographically different forms, a and b, of the enzyme and its precursor are known and a third form seems likely (see bovychymosin, bovychymoc). This sequence has tentatively been identified as coding for preprochymosin b. Sequence comparison indicates that the precursors for prechymosin a and b differ by only two amino acids, and for b and c by only four amino acids.
 FEATURES
 source Location/Qualifiers
 1..1305
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 <1..1305
 /product="chymosin mRNA"
 26..1171
 /note="preprochymosin b"
 /codon_start=1
 /protein_id="AAA30448.1"
 /db_xref="GI:162860"
 /translation="MRCIVLLAVPALSGAFIRIPLYKSKSLRKALKERHLEDFL QKQDGISSKYSKSGFEVAVSPVTNYLDSQYFGKILYLPPEFTVLPTGSSDFWPS IYCKSNACKNORPDPRKSTFONLGRPLSHYGSGMOGILGDTVLSNVLQOT VGLSTQBPQDVFYVAEPDGLGMAVPSLASVSLPVPNNMNRHLVADLPSVMDRN GOESMLTGLAINPSYTGSLHWPVVOQYMOFDTVTISGVVACGCGCOALIDTG TSKVGPSSDILNIOATGATONOVGEEDIDCNDLSNWPVFRINGMYPLTPSAYT SODQFCTSGFSGFSENHSGKILGDVFRIRYTSVPRANNVLGAKAT"
 29..73
 /note="presequence"
 mat_peptide 200..1168
 /product="chymosin b"
 BASE COUNT 305 a 393 c 340 g 267 t
 ORIGIN
 pst-1 site.
 Query Match 76.8%; Score 900.4; DB 4; Length 1305;
 Best Local Similarity 87.8%; Pred. No. 1.3e-265;
 Matches 982; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
 QY 56 TCGTTCCTTACTACAGCTGTGATGATCAACCGATCTCTCTCAACAAAGTATGCTC 115
 Db 54 TCTTGCTCTCTCCCAAGCGCTGATGATCAACGATCTCTCTCAACAAAGGCAATCTC 113
 QY 116 TCCGTAAGGCGCTGAAGAACATGACTTCTAGAAGACTTCTTGAGAAACAAGATAG 175

114 TGAAGAGGCGCTGAGAGACATGGCTTCTTGAGGACCTTCGACAGAAACAGCAGTATG 173
176 GCATCAGCAGCAAGTACTCCGGCTTCGGTGAAGTTGCTAGCGTCCACTTACCACTACC 235
174 GCATCAGCAGCAAGTACTCCGGCTTCGGGAGGAGGCGCAGCGTGGCCCTGACCACTACC 233
236 TTGATAGTCAATCTTTGGGAGATCTACTCCGAAACCCCGCTCAAGAGTTGACCGTTTC 295
234 TGATAGTCAATCTTTGGGAGATCTACTCCGAAACCCCGCTCAAGAGTTGACCGTTTC 293
296 TCTTTGATAGTCAATCTTTGGGAGATCTACTCCGAAACCCCGCTCAAGAGTTGACCGTTTC 355
294 TGTTTGACATGCTGCTCTGACTTCTGGGTAACCTTATCTAGTGAAGAGATGCTT 353
356 GCAAGAACCAACCAAGTTGATCCGAGAAAGTGTCCACCTTCCAGAACTTAGGCAAAC 415
354 GCAAAACCAACCAAGCTTGCAGCCGAGAAAGTGTCCACCTTCCAGAACTTAGGCAAGC 413
416 CTTTGTCTATCACTAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 475
414 CTTTGTCTATCACTAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 473
476 CTGTCTCAACATTTGGAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 535
474 CTGTCTCAACATTTGGAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 533
536 ATGTCTTCACTATGAGAAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 595
534 ACCTCTTCACTATGAGAAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 593
596 CAGAGTCTGATACCTGTGTTTGAACAATGATGAACCGACACTTATGATGATGATGATGATG 655
594 CAGAGTCTGATACCTGTGTTTGAACAATGATGAACCGACACTTATGATGATGATGATGATG 653
656 TGTCTCGGTTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 715
654 TGTCTCGGTTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 713
716 ATCCATCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 775
714 ACCGCTCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 773
776 AATTCACTGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 835
774 AGTTCACTGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 833
836 AAGCTATCTGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 895
834 AAGCTATCTGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 893
896 TTAGCAGAGCTATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 955
894 TTAGCAGAGCTATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 953
956 ACCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 1015
954 ACCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 1013
1016 CTTTGTCTATCACTAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1075
1014 CTTTGTCTATCACTAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1073
1076 AATTCCAGAAATGATCTTGGAGATGTTTCACTGATGATGATGATGATGATGATGATGATG 1135
1074 AATTCCAGAAATGATCTTGGAGATGTTTCACTGATGATGATGATGATGATGATGATGATG 1133
1136 GGGGCAACCACTGTTGGGCTAGCTTAAAGCATCTGA 1173
1134 GGGGCAACCACTGTTGGGCTAGCTTAAAGCATCTGA 1171

RESULT 4

AR002347
LOCUS AR002347 1240 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 2 from patent US 5741665.
ACCESSION AR002347
VERSION AR002347.1 GI:3963901
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1240)
AUTHORS Kato, E. K. and Stuart, M. Dorsey.
TITLE Light-regulated promoters for production of heterologous proteins
in filamentous fungi
JOURNAL Patent: US 5741665-A 2 21-APR-1998;
FEATURES
source location/Qualifiers
BASE COUNT 274 a 374 c 339 g 253 t
ORIGIN
Query Match 76.5%; Score 897.2; DB 6; Length 1240;
Best Local Similarity 87.7%; Pred. No. 1.3e-264;
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 56 TCGTGTCTGTTACTGACGCTGCTGAGATGACCCGGCATTCCTCTTCAAGAGTAACTTC 115
DB 100 TCTTGTCTCTCTCCAGGCGCTGAGATGACAGGATCCCTCTGTACAAAGGCAAGTCTC 159
QY 116 TCCGTAAGGCGCTGAGAGAAATGGAATCTTGAAGAACTTCTTGCAGAAACACAGATAG 175
DB 160 TGAAGAGGCGCTGAGAGAGATGAGGCTTCTGAGAGATCTTCCGAGAAACAGCAGTATG 219
QY 176 GCATCAGCAGCAAGTACTCCGGCTTCGGTGAAGTTGATGATGATGATGATGATGATGATG 235
DB 220 GCATCAGCAGCAAGTACTCCGGCTTCGGGAGGATGAGCAGCGTCCCTGACCAACTACC 279
QY 236 TTGATAGTCAATCTTTGGGAGATGCTCAACCTTGAAGCTATGATGATGATGATGATGATG 295
DB 280 TGATAGTCAATCTTTGGGAGATGCTCAACCTTGAAGCTATGATGATGATGATGATGATG 339
QY 296 TCTTTGATAGTCAATCTTTGGGAGATGCTCAACCTTGAAGCTATGATGATGATGATGATG 355
DB 340 TGTTTGACATGCTGCTCTGACTTCTGGGTAACCTTATCTAGTGAAGAGATGCTT 399
QY 356 GCAAGAACCAACCAAGTTGATCCGAGAAAGTGTCCACCTTCCAGAACTTAGGCAAAC 415
DB 400 GCAAAACCAACCAAGCTTGCAGCCGAGAAAGTGTCCACCTTCCAGAACTTAGGCAAGC 459
QY 416 CTTTGTCTATCACTAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 475
DB 460 CTTTGTCTATCACTAGGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 519
QY 476 CTGTCTCAACATTTGGAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 535
DB 520 CTGTCTCAACATTTGGAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 579
QY 536 ATGTCTTCACTATGAGAAATTCAGATTCAGACAGATGAGGACTTGAACCCCAAGAACAGGTG 595
DB 580 ACCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 639
QY 596 CAGAGTCTGATACCTGTGTTTGAACAATGATGAACCGACACTTATGATGATGATGATGATG 655
DB 640 CAGAGTCTGATACCTGTGTTTGAACAATGATGAACCGACACTTATGATGATGATGATGATG 699
QY 656 TGTCTCGGTTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 715
DB 700 TGTCTCGGTTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 759
QY 716 ATCCATCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 775
DB 760 ACCGCTCTTACATGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATG 819
QY 776 AATTCACTGAGCAGGAAATGCGCAGAGAGATGCTCAACCTTGAAGCTATGATGATGATGATG 835

[illegible]

FEATURES	source	Location/Qualifiers	FT	mat_peptide	278..1347
BASE COUNT		328 a 440 c 398 g 294 t			
ORIGIN					
Query Match		76.5%; Score 897.2; DB 6; Length 1460;			
Best Local Similarity		87.7%; Pred. No. 1.3e-264;			
Matches	980; Conservative	0; Mismatches 138; Indels 0; Gaps 0;			
Db					
56	TCGTGCTGTACTACACGCTGCTAGATCACCCGCATTCCTCTTACAAAGTAAGTCTC	115			
233	TCTTCGCTCTCTCCAGAGGCGCTGAGATCACAGAGATCCCTGTGACAAAGCAAGTCTC	292			
Qy	116	TCGTAAGGCGCTGAAGAACATGAGACTTCTTGAAGACTTTTGCAGAAACACAGATG	175		
Db	293	TGAAGAAAGCGCTGAAGAGATGAGGCTTTCGAGAGACTTCTCAGAAACAGAGATG	352		
Qy	176	GCATCAGACAGAGTACCTCCGCTTCGCTGTAAGTTGCTAGAGGTGCACTTACCAATAC	235		
Db	353	GCATCAGACAGAGTACCTCCGCTTCGCTGTAAGTTGCTAGAGGTGCACTTACCAATAC	412		
Qy	236	TTGATAGTACATATTGGGAAGATCTACCTCGAAACCCGCGCTCAAGAGTTCAACCGTTC	295		
Db	413	TGATAGTACATATTGGGAAGATCTACCTCGAAACCCGCGCTCAAGAGTTCAACCGTTC	472		
Qy	296	TCTTTGATCTGCTTCTCTGACATTTCTGGGTTCCCTCTATTTACTGCAAGACAAATGCT	355		
Db	473	TGTTTGACATGCTGCTCTGACATTTCTGGGTTCCCTCTATTTACTGCAAGACAAATGCT	532		
Qy	356	GCAAGAACCAACCAAGATTCGATCCGAAAGTGTCCACCTTCAGAACTTAGCAAC	415		
Db	533	GCAAGAACCAACCAAGATTCGATCCGAAAGTGTCCACCTTCAGAACTTAGCAAC	592		
Qy	416	CCCTGTCTATCTACCTAGCTGACAGTGAACATGCAAGATCTTAGGCTATGATACCGTCA	475		
Db	593	CCCTGTCTATCTACCTAGCTGACAGTGAACATGCAAGATCTTAGGCTATGATACCGTCA	652		
Qy	476	CTGTCTCCAACTTGTGACATTTCAACAGACAGTAGCACTTAGCAACCAAGACAGGTG	535		
Db	653	CTGTCTCCAACTTGTGACATTTCAACAGACAGTAGCACTTAGCAACCAAGACAGGTG	712		
Qy	536	ATGTCTTCACTTAAGCAAAATTCATGAGCATCCCTTGGATATGACATCCCATCCGCTCGGT	595		
Db	713	ACGTCTTCACTTAAGCAAAATTCATGAGCATCCCTTGGATATGACATCCCATCCGCTCGGT	772		
Qy	596	CAGAGTACTCGATACCTGTGTTTGAACCAATGATGAACCGACACTAGTACTCGAAGCT	655		
Db	773	CAGAGTACTCGATACCTGTGTTTGAACCAATGATGAACCGACACTAGTACTCGAAGCT	832		
Qy	656	TGTTCTCGGTTTACATGACAGAGATGGCCAGAGAGCATCTCAACGCTTGAAGCTATTG	715		
Db	833	TGTTCTCGGTTTACATGACAGAGATGGCCAGAGAGCATCTCAACGCTTGAAGCTATTG	892		
Qy	716	ATCCATCTTACATACAGAGATCTCTTCACTGAGGTTCCAGTACTGTGTGACAGATCTG	775		
Db	893	ATCCATCTTACATACAGAGATCTCTTCACTGAGGTTCCAGTACTGTGTGACAGATCTG	952		
Qy	776	AATTCACGTGGAACAGTACACATACAGCGGTGTGTTGCAATGTAAGTGAATGTC	835		
Db	953	AGTTCACTGTGGAACAGTACACATACAGCGGTGTGTTGCAATGTAAGTGAATGTC	1012		
Qy	836	AAGCTATCTTGAATACCGGTACGTCCAAAGCTGTGCGACCTTAGACGCACTTCTCAACA	895		
Db	1013	AAGCTATCTTGAATACCGGTACGTCCAAAGCTGTGCGACCTTAGACGCACTTCTCAACA	1072		
Qy	896	TTTCAGCAAGCTATTGAGGCCACACAGAACAGTACCGGTATTTGACATATGATTTGACA	955		
Db	1073	TTTCAGCAAGCTATTGAGGCCACACAGAACAGTACCGGTATTTGACATATGATTTGACA	1132		

Query Match 76.4%; Score 895.6; DB 6; Length 1175;
Best Local Similarity 87.6%; Pred. No. 4e-264;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 56 TCGTTGCTGTTACTACGCTGCTGATGATCCCGCATTCCTCTTACAAAGGTAAGTCTC 115
DB 49 TCTTGTCTCTCTCCCGAGGCGCTGAGATCACAGGATCCTCTGTAAAGCAAGTCTC 108
QY 116 TCCGTAAGCGCTGAAGAACATGAGCTTTAGAGACTCTCTGAGAAACAACAGTATG 175
DB 109 TGAGAGAGCGCTGAGAGAGCATGAGGCTTCTGAGAGACTCTCTGAGAAACAGCATATG 168
QY 176 GCATCAGCAGCAAGTACTCCGCTTCGCTGAGTTCAGTCTGACGTGCACTTACCACTAC 235
DB 169 GCATCAGCAGCAAGTACTCCGCTTCGCTGAGTTCAGTCTGACGTGCACTTACCACTAC 228
QY 236 TTGATGATCAATCTTTGAGAAAGATCTACCTCGAAACCCCGCTCAAGATTACCGTTG 295
DB 229 TGGATGATCAATCTTTGAGAAAGATCTACCTCGAAACCCCGCTCAAGATTACCGTTG 288
QY 296 TCTTTGATACGTGTTCTCTGATCTTCTGAGTTCCTCTATCTACTGCAAGGCAATGCT 355
DB 289 TGTGTGACATGCGCTCTCTGATCTTCTGAGTTCCTCTATCTACTGCAAGGCAATGCT 348
QY 356 GCAAGAACCAACCAAGATTGATCCGAGAAAGTCTGCACTTCAGAACTTACGGCAAC 415
DB 349 GCAAGAACCAACCAAGATTGATCCGAGAAAGTCTGCACTTCAGAACTTACGGCAAC 408
QY 416 CTTGTCTATACACTACGCTGATCAGGTAGCATGCAAGAACTTATGATGATGATCTCA 475
DB 409 CCCTGTCTATCACTACGCTGATCAGGTAGCATGCAAGAACTTATGATGATGATCTCA 468
QY 476 CTGTCTCAATGTTGATGATCTTCAACAGCATGAGTTCAGTTCAGTTCAGTTCAGTTC 535
DB 469 CTGTCTCAATGTTGATGATCTTCAACAGCATGAGTTCAGTTCAGTTCAGTTCAGTTC 528
QY 536 ATGTCTTACCTATGATCAGAAATTCGATGATCTTCTGATGATGATGATGATGATGAT 595
DB 529 ACGTCTTACCTATGATCAGAAATTCGATGATCTTCTGATGATGATGATGATGATGAT 588
QY 596 CAGAGTACTGATACCTGTGTTGACCAATGATGAACCGACCTTATGATGATGATGAT 655
DB 589 CAGAGTACTGATACCTGTGTTGACCAATGATGAACCGACCTTATGATGATGATGAT 648
QY 656 TGTTCCTGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 715
DB 649 TGTTCCTGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
QY 716 ATTCATCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 775
DB 709 ACCGTCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
QY 776 AATTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 835
DB 769 AGTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 828
QY 836 AAGCTATCTTGGATACCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 895
DB 829 AAGCTATCTTGGATACCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 888
QY 896 TTTCAGCAGCATTTGAGAGCAGACAGAACCATGATGATGATGATGATGATGATGATGAT 955
DB 889 TCCAGCAGCATTTGAGAGCAGACAGAACCATGATGATGATGATGATGATGATGATGAT 948
QY 956 ACCTTAGTACATGCTCAGTGTCTTTGAGATGATGATGATGATGATGATGATGATGATGAT 1015
DB 949 ACCTTAGTACATGCTCAGTGTCTTTGAGATGATGATGATGATGATGATGATGATGATGAT 1008
QY 1016 CCTCGGCTTATACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1075
DB 1009 CCTCGGCTTATACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1068

QY 1076 ATTCCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1135
DB 1069 ATTCCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1128
QY 1136 GGGCCAAACCTGTTGGGCTAGTAAAGCAATGTA 1173
DB 1129 GGGCCAAACCTGTTGGGCTAGTAAAGCAATGTA 1166

RESULT 8
AR073077 2733 bp DNA linear PAT 28-AUG-2000
LOCUS Sequence 6 from patent US 5948682.
DEFINITION AR073077
ACCESSION AR073077
VERSION AR073077.1 GI:9999840
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2733)
AUTHORS Moloney, M. M.
TITLE Preparation of heterologous proteins on oil bodies
JOURNAL Patent: US 5948682-A 6 07-SEP-1999;
FEATURES Location/Qualifiers
source 1..2733
BASE COUNT 699 a 681 c 684 g 669 t
ORIGIN

Query Match 76.3%; Score 894.8; DB 6; Length 2733;
Best Local Similarity 88.4%; Pred. No. 7.6e-264;
Matches 971; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 76 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 135
DB 1630 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1689
QY 136 CATGACCTTCTAGAAAGCTTCTGAGAAACAGATGATGATGATGATGATGATGATGATGAT 195
DB 1690 CATGACCTTCTAGAAAGCTTCTGAGAAACAGATGATGATGATGATGATGATGATGAT 1749
QY 196 GCGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255
DB 1750 GCGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1809
QY 256 AAGATCTACCTGGAACCCCGCTTAAAGTTCACCTGTTCTTGTGATGATGATGATGAT 315
DB 1810 AAGATCTACCTGGAACCCCGCTTAAAGTTCACCTGTTCTTGTGATGATGATGATGAT 1869
QY 316 GACTTCGAGGTTCCCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
DB 1870 GACTTCGAGGTTCCCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
QY 376 GATCCGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435
DB 1930 GATCCGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1989
QY 436 AAGGTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 495
DB 1990 AAGGTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2049
QY 496 ATTCAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
DB 2050 ATTCAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2109
QY 556 TTTCAGCAGCATTTGAGAGCAGACAGAACCATGATGATGATGATGATGATGATGATGAT 615
DB 2110 TTTCAGCAGCATTTGAGAGCAGACAGAACCATGATGATGATGATGATGATGATGATGAT 2169
QY 616 TTTGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 675
DB 2170 TTTGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2229

QY	676	AGGAATGGCCAGAGAGAGATCTCAGCTTGGAGCTATTGATTCATCTACTACACAGAGA	735
Db	2230	AGGAATGGCCAGAGAGAGATCTCAGCTTGGAGCTATTGATTCATCTACTACACAGAGA	2289
QY	736	TCTTCTCACTGGGATTCAGTCACTGTGCAGCAGTACTGGCAATTCACTGTGGACAGTGTG	795
Db	2290	TCTTCTCACTGGGATTCAGTCACTGTGCAGCAGTACTGGCAATTCACTGTGGACAGTGTG	2349
QY	796	ACCATCAGCGGTGTGGTGTGTGTCATGTGAAGTGGATGTCAAGCTATTTTGGATACCGGT	855
Db	2350	ACCATCAGCGGTGTGGTGTGTGTCATGTGAAGTGGATGTCAAGCTATTTTGGATACCGGT	2409
QY	856	ACGTCGAAGCTGGTGGGACCTTAGACGAGCAATTCGCAATTCGACAGTATTGGAGACC	915
Db	2410	ACGTCGAAGCTGGTGGGACCTTAGACGAGCAATTCGCAATTCGACAGTATTGGAGACC	2469
QY	916	ACACAGAACCCAGTACGGGTGATTTGACATTAATTCGCAACACCTTAGCTACATGCCCTACA	975
Db	2470	ACACAGAACCCAGTACGGGTGATTTGACATTAATTCGCAACACCTTAGCTACATGCCCTACA	2529
QY	976	GTTGTCTTTTGAGATCAACGCGCAAGTGTACCCACTGACCCCTCCGCTTATACAGCCAG	1035
Db	2530	GTTGTCTTTTGAGATCAACGCGCAAGTGTACCCACTGACCCCTCCGCTTATACAGCCAG	2589
QY	1036	GATCAAGGGGTTCTGCACACCGAGTAAATTCAGAGTGAGAACCAATTCGCAAAATGATCTGT	1095
Db	2590	GATCAAGGGGTTCTGCACACCGAGTAAATTCAGAGTGAGAACCAATTCGCAAAATGATCTGT	2649
QY	1096	GGAGATGTGTTCAATTCGTGAGTACTACAGCGTCTTTTGA CAGGGCCAA CAACTCGTTGGG	1155
Db	2650	GGAGATGTGTTCAATTCGTGAGTACTACAGCGTCTTTTGA CAGGGCCAA CAACTCGTTGGG	2709
QY	1156	CTAGCTTAAGCAATCTGA 1173	
Db	2710	CTAGCTTAAGCAATCTGA 1173	
RESULT 9			
BOVCHYMOSA			
LOCUS	1275 bp	mRNA	linear
DEFINITION	Bos taurus preprochymosin a mRNA, complete cds.		MAM 19-DEC-2002
ACCESSION	J00002		
VERSION	J00002.1	GI:162857	
KEYWORDS			
SOURCE	Bos taurus (cow)		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos. 1 (Bases 1 to 1275)		
AUTHORS	Moir, D., Mao, J., Schumm, J. W., Vovis, G. F., Alford, B. L. and Taunton-Rigby, A.		
TITLE	Molecular cloning and characterization of double-stranded cDNA coding for bovine chymosin		
JOURNAL	Gene 19 (1), 127-138 (1982)		
MEDLINE	83054629		
PUBMED	6183168		
COMMENT	Chymosin (rennin) is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically distinct forms are known and a third seems likely (see bovchymob and bovchymoc). this sequence has been tentatively identified as chymosin a; it differs from chymosin b by only two amino acids and from chymosin c by another amino acid. [1] argues that the different chymosins are probably polymorphic variants of a single chymosin gene.		
FEATURES	Location/Qualifiers		
source	1..1275	/organism="Bos taurus"	
	/mol_type="mRNA"		
	/db_xref="taxon:9913"		
	/feature_type="fourth stomach mucosa"		
CDS	21..1166	/note="rennin"	

	sig_peptide	mac_peptide	BASE COUNT	293 a	391 c	336 g	255 t	ORIGIN
								/codon_start=1 /product="preprochymosin a" /protein_id="AAA0447.1" /db_xref="GI:162858" /translation="MRLVLLAVPALSOGAEITRIPLKYSKSLRKALKHEGLDFLQKQYGISKSYSGEFAVSPFLINYLDSQYSGKLYLGPPOEFTVLPDTSGLSDFWPPIKYSNACKNHQRDFPRKSPFONMGLDLSHYGSGNOGILGDIYVSNLIDLQOQVGLSTQEPDPEFYAEFDGILGMAVPLASYSYSLPVPNNMNRHLVADLPSVWDRRGESMLTLGALIDPCYTGSGTHWVYVQOIFPVDSTVYSIGVVACGCAALIDTSTKSLGPSSDILNIQOALIGATONQYDEFDIDCNLSIPTYTVFLINMGYPLTPSAVYSODGFCSTSGFQSGENHSQKWLIGDVIFREYYSVPDRANNLVGLAKAI" 24..68 /note="presequence"
								Query Match 76.2%; Score 894; DB 4; Length 1275; Best Local Similarity 87.5%; Pred. No. 1,3e-263; Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0
Qy	56	TCGTTGCGTTACTACACGCTGCTGAGATCACCCGCACTCTCTACAAAGTAAAGTCTC	115					
Db	49	TCTTGCTCTCTCCAGAGGCGCTAGATCACAGAGATCCCTCTGTACAAAGCAGTCTC	108					
Qy	116	TCGCTAAGCGCCTGAAGAACATGAGACTTCTTAAGAACTTCTTGCAAAACACAGTATG	175					
Db	109	TGAGAAAGCGCCTGAAGAGCATGGGCTTCTGAGAGACTTCTCTGCAAAACAGCAGTATG	168					
Qy	176	GCATCAGAGAGAAATCTCCGGCTTCGGGTGAAGTTGTAAGGAGTGCATTTACCACTAAC	235					
Db	169	GCATCAGAGAGAAATCTCCGGCTTCGGGTGAAGTTGTAAGGAGTGCATTTACCACTAAC	228					
Qy	236	TTGATAGCAATPACTTTGGGAAGATCTACCTCGGAACCCGCGCTAAGAGTTCAACCTTC	295					
Db	229	TGATAGTCAATPACTTTGGGAAGATCTACCTCGGAACCCGCGCTAAGAGTTCAACCTTC	288					
Qy	296	TCCTTGATCTGGTTCCTCTGACTTCTGGGTTCCCTCTATCTATCTGAAGAGAAATGCCCT	355					
Db	289	TGTTTGACACTGGCTCCTCTGACTTCTGGGTACCTCTATCTATCTGAAGAGCAATGCCCT	348					
Qy	356	GCAAGAACCAACCAAGATTGCATCCGAGAAAGTGCCTCACTTCCAGAACTTAGGCAAC	415					
Db	349	GCAAAACCAACCAAGCTTCGACCCCGAAGAAAGTGCCTCACTTCCAGAACCTGGGCAAGC	408					
Qy	416	CCTTGCTATACATACGCTACAGGTACATGACAGAACTTTAGGCTATGATGATCGTCA	475					
Db	409	CCCTGCTATACATACGAGACAGGACAGCATGACAGGAGATCTGGGCTATGATGACCTGCA	468					
Qy	476	CTGTCTCCMACTTGTGTGACATTGACACAGACAGTAGAATTAGCAACCAAGAACAGAGTG	535					
Db	469	CTGTCTCCMACTTGTGTGACATTGACACAGACAGTAGAGCTAGACACCAAGAGGCCGGGG	528					
Qy	536	ATGTCTTCACTATGACGAATTCATGAGCATTCCTTGGTATGGCATTAACCATGCTTCGGCT	595					
Db	529	ACGTCTTCACTATGACGAATTCATGAGCATTCCTTGGTATGGCATTAACCATGCTTCGGCT	588					
Qy	596	CAGAGTACTGATPACCTGTGTTTACACATGATGAAACCGACACCTTAGTATGCTCAAGACT	655					
Db	589	CAGAGTACTGATPACCTGTGTTTACACATGATGAAACCGACACCTTAGTATGCTCAAGACC	648					
Qy	656	TGTTCTCGGTTTAAATGACAGAAATGAGCAGAGAGAGCATCTCACGCTTGGAGCTATTTG	715					
Db	649	TGTTCTCGGTTTAAATGACAGAAATGAGCAGAGAGAGCATCTCACGCTGAGGAGCATTTG	708					
Qy	716	ATCATCTCTAATCAACAGAGATCTCTTCACTGAGGTTCCAGTCACTGTGACAGCATGATGCGC	775					
Db	709	ACCGTGTCTAATCAACAGAGATCTCTTCACTGAGGTTCCAGTCACTGTGACAGCATGATGCGC	768					
Qy	776	AATTCACCTGTGACAGTGTCAACATCAAGCGGTGTGTGTTGCAATGTGAAGTGGATGTGC	835					
Db	769	AGTTCACTGTGACAGTGTCAACATCAAGCGGTGTGTGTTGCAATGTGAAGTGGATGTGC	828					

QY 836 AAGTATCTTGATACCGGATCGTCAAGCTGGTGGACCTAGCAGCCGACATCTCAACA 895
 DB 829 AGGCATCTGGACACGGGACACCTCCAGCTGTGGGGCCGACGACGACATCTCCAAAC 888
 QY 886 TTCAGCAAGTATTGGAGCCACACAGAACCAAGTACGGTGAATTGATAGATTGCCGACA 955
 DB 889 TCCAGCAGGCGCATTTGGAGCCACACAGAACCAAGTACGATGATTTGATCATCGACTCGACA 948
 QY 956 ACCTAGTACATGCTACAGTGTCTTGAGATCAAGGCAAGATGATACCACTGACCC 1015
 DB 949 ACCTAGTACATGCTACAGTGTCTTGAGATCAAGGCAAGATGATACCACTGACCC 1008
 QY 1016 CCTCGCCTTATACAGCAGGATCAAGGATCTGACAGGATTTCCAGAGTGAACC 1075
 DB 1009 CCTCGCCTTATACAGCAGGATCAAGGATCTGACAGGATTTCCAGAGTGAACC 1068
 QY 1076 ATTCGCAAGATGATCTTGGAGATGTTCATTCTGATGATCTACAGCGTCTTTGACA 1135
 DB 1069 ATTCGCAAGATGATCTTGGAGATGTTCATTCTGATGATCTACAGCGTCTTTGACA 1128
 QY 1136 GGGCCAAACACTCTTGGGCTAGCTAAAGCAATCTGA 1173
 DB 1129 GGGCCAAACACTCTTGGGCTAGCTAAAGCAATCTGA 1166
 RESULT 10
 LOCUS E00075 1290 bp RNA linear PAT 29-SEP-1997
 DEFINITION E00075 cDNA encoding chymosin (=rennin).
 ACCESSION E00075
 VERSION E00075.1 GI:2168379
 KEYWORDS JP 1983009687-A/1.
 SOURCE Bos taurus (cow).
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 1290)
 Noeman,H.K., Maikere,T.D., Chimoshi,J.R.H.H., Pilita,A.R. and
 Jiyon,S.E.
 TITLE PRODUCTION OF POLYPEPTIDE
 JOURNAL Patent: JP 1983009687-A 1 20-JAN-1983;
 CELLTECH LTD
 COMMENT OS bovine
 PN JP 1983009687-A/1
 PD 20-JAN-1983
 PR 17-JUN-1982 JP 1982104672
 01-DEC-1981 GB 81 8136185, 10-FEB-1982 GB 82 8203907 PI
 NOOMAN HERRI KEARI, MAIKERU TERENSU DOORU,
 PI CHIMOSHII JIYON KOI HARISU,
 PII PIIITA ANSONIT ROU, JIYON SUPENSNA EMUTEBUJI
 PC C12N9/52,C07H21/04,C12N1/00,C12N9/60,C12N15/00, PC
 C12Q1/02//C12R1/19,
 PC C12R1/865;
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC *source: tissue type=stomach;
 CC *source: clone=pCT 70;
 FH key Location/Qualifiers
 FT 5'UTR 1..25
 FT sig_peptide 26..73
 FT mat_peptide 74..1078
 FT /product='pro-chymosin',
 FT 3'UTR 1082..1200
 FT CDS 26..1081
 FT /product='pre-pro-chymosin',
 Location/Qualifiers
 1..1290
 source /organism="Bos taurus"

BASE COUNT 309 a 384 c 334 g 263 t
 ORIGIN
 Query Match 76.2%; Score 894; DB 6; Length 1290;
 Best Local Similarity 87.5%; Pred.No. 1.3e-263;
 Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
 /mol_type="genomic RNA"
 /db_xref="taxon:9913"
 56 TCGTGTCTGTTACTACAGCTCTGAGATCACCCGATTCCTCTTACAAAGTAAGTCTC 115
 DB 54 TCTTGTCTCTCTCCCAAGGCGCTGATCACAGATCTCTGTACAAAGCAAGTCTC 113
 QY 116 TCCGTAAGCGCGTGAAGGAATGAGACTCTTGAAGACTCTTGAGAAACAAGATG 175
 DB 114 TGAGGAAGCGCGTGAAGGAATGAGACTCTTGAAGACTCTTGAGAAACAAGATG 173
 QY 176 GCATCAGCAGCAAGTACTCCGCTTGGTGAAGTTGCTAGCGTGCACCTTACCACTACC 235
 DB 174 GCATCAGCAGCAAGTACTCCGCTTGGGGAAGTGCAGAGGTGCCCTGACCAACTACC 233
 QY 236 TTGATAGTCAATCTTTGGGAAGATCTACCTCGGAACCCCGCTTGAAGTTCACCGTTC 295
 DB 234 TAGATAGTCAATCTTTGGGAAGATCTACCTCGGAACCCCGCTTGAAGTTCACCGTTC 293
 QY 296 TCTTGTATAGTGTCTCTCTGATCTTGGGTTCCCTTATCTACTGCAAGAGCAATGCCCT 355
 DB 294 TGTGTGACACTGGCTCTCTCTGATCTTGGGTTCCCTTATCTACTGCAAGAGCAATGCCCT 353
 QY 356 GCAAGAACCAACCAAGATTCGATCCGAAAGTCTCCACCTTCCAGAACTTGAAGCAAC 415
 DB 354 GCAAGAACCAACCAAGATTCGATCCGAAAGTCTCCACCTTCCAGAACTTGAAGCAAC 413
 QY 416 CTTGTCTTATACACTACGCTACAGGTACATGCAAGAACTTTAGGCTATGATACCGTCA 475
 DB 414 CCTGTCTTATACACTACGCTACAGGTACATGCAAGAACTTTAGGCTATGATACCGTCA 473
 QY 476 CTGTCTTCAAAATGTGTGAACATTCGACAGAGTGAAGCTTGAAGCAAGCAAGAGT 535
 DB 474 CTGTCTTCAAAATGTGTGAACATTCGACAGAGTGAAGCTTGAAGCAAGCAAGAGT 533
 QY 536 AATGCTTCACTTATGCAAGATTCGATGAGCTCTTGTATGAGCATACCATGCTCGCT 595
 DB 534 AATGCTTCACTTATGCAAGATTCGATGAGCTCTTGTATGAGCATACCATGCTCGCT 593
 QY 596 CAGAGTACTGATACCTGTGTTTGAACAATGATGAACCGACACTAGTACTCAAGACT 655
 DB 594 CAGAGTACTGATACCTGTGTTTGAACAATGATGAACCGACACTAGTACTCAAGACT 653
 QY 656 TGTCTCTCGGTTTACATGACAGGAATGCGCAGAGAGATGTCACCGCTTGAAGCTATTG 715
 DB 654 TGTCTCTCGGTTTACATGACAGGAATGCGCAGAGAGATGTCACCGCTTGAAGCTATTG 713
 QY 716 ATTCATCTTACTACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGAGTACTG 775
 DB 714 ATTCATCTTACTACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGAGTACTG 773
 QY 776 AATTCATCTGGAACAGTGTCAACATCAGCGTGTGTGTTGTCATGTGAAGGTGATGTC 835
 DB 774 AATTCATCTGGAACAGTGTCAACATCAGCGTGTGTGTTGTCATGTGAAGGTGATGTC 833
 QY 836 AAGCTATCTTGGATACCGGTACAGTGTGTGAGCTGTGAGACTTGAAGCAAGTCTTCAACA 895
 DB 834 AAGCTATCTTGGATACCGGTACAGTGTGTGAGCTGTGAGACTTGAAGCAAGTCTTCAACA 893
 QY 896 TTCAGCAAGTATTGGAGCCACACAGAACCAAGTACGGTGAATTGATAGATTGCCGACA 955
 DB 894 TCCAGCAGGCGCATTTGGAGCCACACAGAACCAAGTACGATGATTTGATCATCGACTCGACA 953
 QY 956 ACCTAGTACATGCTACAGTGTCTTGAGATCAAGGCAAGATGATACCACTGACCC 1015
 DB 954 ACCTAGTACATGCTACAGTGTCTTGAGATCAAGGCAAGATGATACCACTGACCC 1013

AUTHORS Chiyaarunu, E. B.
 TITLE CATTLE KIMOCIN
 JOURNAL Patent: JP 1984021392-A 1 03-FEB-1984;
 GENEX CORP
 COMMENT OS bovine calf
 PN JP 1984021392-A/1
 PD 03-FEB-1984
 PF 30-JUN-1983 JP 1983119481
 PR 01-JUL-1982 US 82 394433, 13-APR-1983 US 83 484539 PI
 CHIYAARUNU ET BASURENSUNO
 PC C12N15/00, C07H21/04, C12N1/20, C12P19/34, C12P21/02, (C12N15/00,
 PC C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: class: type=stomach mucosa;
 FH Location/Qualifiers
 FT mat_peptide 169..1113
 FT FT /product='bovine calf chymosin' FT CDS
 FT <1..1116 /product='precursor polypeptide'.
 FEATURES
 source Location/Qualifiers
 1..1289
 /organism="unidentified"
 /mol_type="genomic RNA"
 /db_xref="taxon:32644"
 BASE COUNT 311 a 388 c 329 g 260 t 1 others
 ORIGIN
 Query Match 76.1%; Score 893; DB 6; Length 1289;
 Best Local Similarity 87.4%; Pred. No. 2.6e-263;
 Matches 977; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
 QY 56 TCGTGTCTGTTACTACGCTGCTGAGATCACCAGATTCTCTCTACAAAGGTAAGTCTC 115
 DB 23 TCTTGCTCTCTCCAGGGGCGCTGAGATCACCAGATCCCTCTGTCACAAAGGCAAGTCTC 82
 QY 116 TCCGTAAAGCGCTGAAGAACTGAGACTTTGTAAGACTTTCTTGACAAACAGATATG 175
 DB 83 TGAGGAGAGCGCTGAAGAGAGCTGAGCTTCTGAGAGACTTCTGCAAGAAACAGCAGTATG 142
 QY 176 GCATGAGAGAGACTCCGGCTCTGGTAAAGTGTAGAGGTGACCTTACCACTACC 235
 DB 143 GCATGAGAGAGACTCTCGGCTCTGGAGGTGACAGCGTCCCTGACCAACTACC 202
 QY 236 TTGATAGTCAATCTTTGGGAAGATCTACCTCGAAACCCCGCTCAAGATTCAACCGTTTC 295
 DB 203 TAGATAGTCAAGTCTTTGGGAAGATCTACCTCGGAGACCCCGCCAGAGATTCAACCGTTC 262
 QY 296 TCTTTGATAGTCTGCTCTCTGACTTCTGGGTTCCCTTATCTAGTCTGCAAGAGCAATGCTC 355
 DB 263 TGTTTGACACTGGCTCTCTGACTTCTGGGTACCTCAATCTAGTCAAGAGCAATGCTC 322
 QY 356 GCAAAACCAACCAAGATTTCGATCCGAGAAAGTCTTCCACCTTCCAGAACTTGAAGCAAC 415
 DB 333 GCAAAACCAACCAAGCTTTCGATCCGAGAAAGTCTTCCACCTTCCAGAACTTGAAGCAAC 382
 QY 416 CTTTGTCTATACATCAGGTAGTACAGGTAGTACAGAAATCTTAAAGCTATGATACCGTCA 475
 DB 383 CCGTGTCTATACATCAGGTAGTACAGGTAGTACAGAAATCTTAAAGCTATGATACCGTCA 442
 QY 476 CTGTCTCCAACTTTGACATTTCAACAGACAGTAAAGATTAGCAACCCAGAAACCAAGGTG 535
 DB 443 CTGTCTCCAACTTTGACATTTCAACAGACAGTAAAGATTAGCAACCCAGAAACCAAGGTG 502
 QY 536 ATGTCTTACCTATGACAGATTCGATGCTTGTGATGAGCATACCATGCTGCTGCT 595
 DB 503 AGCTTCTACCTATGACAGATTCGATGCTTGTGATGAGCATACCATGCTGCTGCTGCT 562
 QY 596 CAGAGTACTGATACCTGTTTGAACAATGATGAAACGAGCACTAGTACCTCAAGACT 655

DB 563 CAGAGTACTGATACCCGCTGTTTGAACAATGATGAACGAGCACTGTGCGCCAAAGCC 622
 QY 656 TGTTCGCTGTTTAAATGAGCAAGAAATGCGCAGAGAGCATGCTGACGCTTGAGCTATTG 715
 DB 623 TGTTCGCTGTTTAAATGAGCAAGAAATGCGCAGAGAGCATGCTGACGCTTGAGCTATTG 682
 QY 716 ATTCATCTTACTACACAGATCTCTTCACTGGTTCCAGTCACTGTGACAGCATGCTGCG 775
 DB 683 ACCGCTCTTACTACACAGAGTCCCTGCACTGGTCCCGTGCACATGTCAGCATGCTGCG 742
 QY 776 AATTACTGTGAGACAGTGTACCATCAGCGGTGTGTGTTGATGTGAAGTGTGATGTC 835
 DB 743 AGTTCATGTGTGAGACAGTGTACCATCAGCGGTGTGTGTTGATGTGAAGTGTGATGTC 802
 QY 836 AAGCTATCTTGATACCGGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 895
 DB 803 AGGCTATCTGAGACAGCGGACCTTCAAGCTGTGCTGCGCCAGCAGCAGCATCTCAACA 862
 QY 896 TTCAGCAAGCTATTGAGACCAACAGAACAGTACAGTGTGATGATGATGATGATGATGATG 955
 DB 863 TCAGCAGAGCATTTGAGACCAACAGAACAGTACAGTGTGATGATGATGATGATGATGATG 922
 QY 956 ACCTTAGCTATGCTATGCTATGCTGCTTGTGATGATGATGATGATGATGATGATGATG 1015
 DB 923 ACCTTAGCTATGCTATGCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 982
 QY 1016 CCTCGGCTTATGACAGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1075
 DB 983 CCTCGGCTTATGACAGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1042
 QY 1076 ATTCAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1135
 DB 1043 ATTCAGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1102
 QY 1136 GGGCCAAACCTGCTGAGGCTGAGTAAAGCAATCTGA 1173
 DB 1103 GGGCCAAACCTGCTGAGGCTGAGTAAAGCAATCTGA 1140
 RESULT 13
 LOCUS 108097 2726 bp DNA linear PAT 02-DEC-1994
 DEFINITION Sequence 5 from Patent EP 0301669.
 ACCESSION 108097
 VERSION 108097.1 GI:589192
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2726)
 AUTHORS van den Berg, J. A. D. and Brake, A. J. D.
 TITLE DNA constructs containing a Kluveromyces alpha-factor leader
 JOURNAL sequence for directing secretion of heterologous derivatives
 Patent: EP 0301669-A1 5 01-FEB-1989;
 FEATURES
 source Location/Qualifiers
 1..2726
 /organism="unknown"
 BASE COUNT 751 a 654 c 572 g 749 t
 ORIGIN
 Query Match 76.1%; Score 893; DB 6; Length 2726;
 Best Local Similarity 87.5%; Pred. No. 2.7e-263;
 Matches 977; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
 QY 57 CGTTGCTGTACTACAGCTGCTGAGATCACCAGATTCTCTCTACAAAGGTAAGTCTCT 116
 DB 669 CGATGCTTCCATCATATGCTGAGATCACCAGATCCCTCTGTACAAAGGTAAGTCTCT 728
 QY 117 CCGTAAAGCGCTGAAGAACTGACTTCTGAAAGACTTTCTTGACAAACAGACTATG 176
 DB 729 GAGGAAGCGCTGAAGAACTGACTTCTGAAAGACTTTCTTGACAAACAGACTATG 788
 QY 177 CATCAGCAGCAAGTACTCGGCTTCCGTGAAGTGTCTAGCGTGCACCTTACCAACTACT 236

```

Db      789 CATCAGCAGAGTACTCCGGCTTCGGGAGGTGGCCAGGTCCCTCCCTGACCAACTACT 848
Qy      237 TGAATGTAATACTTTGGGAAAGTCTA CTTGGAACCCGCGCTCAAGAGTTCA CCGTTCT 296
Db      849 GGACAGTCACTTTGGGAAAGTACTACTCGGAAACCCGCCACAGAGTTCA CCGTGTCT 908
Qy      297 CTTTGAATACGTGTTCCCTGACCTTGGGGTCCCTCTATCTTACTTGAAGAGCAATGCTG 356
Db      909 GTTTGACACTGGCTCTGACTTCTGGGGTACCCCTATCTTACTTGAAGAGCAATGCTG 968
Qy      357 CAAGAACAACAAGATTGATCCGAGAAAGTGTCCACCTTCAGAACTTAAAGCAAAACC 416
Db      969 CAAAAACCAACAGCGCTTTCGACCCGAGAAAGTGTCTCACTTCCAGAACTGGGCAAGCC 1028
Qy      417 CTTTGTATACACTACGATGATCAGGTAGCATGACAGAAATCTTAAGCTATGATACCGTCA 476
Db      1029 CCGTGTATCACTACGAGGACAGGACATGACAGGAGCATCTGGGCTATGACACCGTCA 1088
Qy      477 TGTCTCAACATTTGTGACATTTCAACAGACATGAGCACTTAAGCAACCAAGACAGGTGA 536
Db      1089 TGTCTCAACATTTGTGACATTTCAACAGACATGAGCACTTAAGCAACCAAGACAGGTGA 1148
Qy      537 TGTCTCACTATGACAAATTCGATGAGCATCTTGTATGACATACCATCGCTCGCGTC 596
Db      1149 CGTCTTCACTATGACCAATTCGATGAGCATCTTGTATGAGCATACCATCGCTCGCGTC 1208
Qy      597 AGAGTACTCGATA CTTGTGTTTGAACAATGATGAAACGACACCTAGTAGCTCAAGATT 656
Db      1209 AGAGTACTCGATACCGTGTGTTGAACAATGATGAAACGACACCTAGTAGCTCAAGATT 1268
Qy      657 GTTCTGGTTTAATGACAGAGAAATGGCCAGAGAGATGCTACCGCTTGGAGCTATGTA 716
Db      1269 GTTCTGGTTTAATGACAGAGAAATGGCCAGAGAGATGCTACCGCTTGGAGCTATGTA 1328
Qy      717 TCCATCTTATACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGGATGAGCA 776
Db      1329 CCGTCTTATACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGGATGAGCA 1388
Qy      777 ATTCACTGTGACAGTGTCAACATCAGCGGTGTGTTGTTGCAATGTGAAAGTGGATGTA 836
Db      1389 GTTCACTGTGACAGTGTCAACATCAGCGGTGTGTTGTTGCAATGTGAAAGTGGATGTA 1448
Qy      837 AGCTATCTTGGATACCGGTACTGTCACAGCTGTGTGACCTTAAGCAGCACTTCTCAAT 896
Db      1449 GGGCATCTTGGATACCGGTACTGTCACAGCTGTGTGACCTTAAGCAGCACTTCTCAAT 1508
Qy      897 TCAGCAAGTATTTGAGGACCAACAAGAACAGTACGATGTTGACATAGATTGCCAACA 956
Db      1509 CAGCAAGGCAATTTGAGGACCAACAAGAACAGTACGATGTTGACATAGATTGCCAACA 1568
Qy      957 CCTTAGCTACATGCTTACAGTTGTCTTTGAGATCAACGGGCAAGATGTACCACTGACCCC 1016
Db      1569 CCGAGCTACATGCTTACAGTTGTCTTTGAGATCAACGGGCAAGATGTACCACTGACCCC 1628
Qy      1017 CTCCGCTTATACCAAGGACGATCAAGGTTCTGACCAAGTGTGATTCAGAGTGAACA 1076
Db      1629 CTCCGCTTATACCAAGGACGATCAAGGTTCTGACCAAGTGTGATTCAGAGTGAACA 1688
Qy      1077 TTCCCAAGAAATGATCTTTGGGAGATGTTGATCTTGTGAGATCTACAGGCTTTTGAAG 1136
Db      1689 TTCCCAAGAAATGATCTTTGGGAGATGTTGATCTTGTGAGATCTACAGGCTTTTGAAG 1748
Qy      1137 GGCACAACACTGTGGGCTAGCTAAAGCAATCTGA 1173
Db      1749 GGCACAACACTGTGGGCTAGCTAAAGCAATCTGA 1785

```

```

RESULT 14
LOCUS   E00108          1311 bp      RNA      linear      PAT 29-SEP-1997
DEFINITION  DNA coding of prepro chymosin.
ACCESSION E00108

```

```

VERSION E00108.1 GI:2168409
KEYWORDS JP 1983109499-A/1.
SOURCE  unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1311)
AUTHORS  Yan.M., Kornueerisu,S.B., Adorianusu,M.R. and Rutsupo,B.
TITLE     RECOMBINED DNA
JOURNAL   Patent: JP 1983109499-A 1 29-JUN-1983;
          UNILEVER NV
          OS      calf
          PN      JP 1983109499-A/1
          PD      29-JUN-1983
          PR      14-OCT-1982 JP 1982180549
          PR      14-OCT-1981 GB 81 8131004
          PI      YAN MAATO, KORNUERISU SEODORUSU BERURITSUPUSU, PI
          ADORIANUSU MARINUSU REDEBERU, RUTSUPO BEBENSU PC
          CC      C07H21/04,C12N1/20,C12N15/00//C12P21/00;
          CC      strandedness: Double;
          CC      topology: Linear;
          CC      hypothetical: No;
          CC      anti-sense: No;
          CC      *source: tissue_location=stomach;
          FH      Key      location/Qualifiers
          FT      CDS      24..1169
                          /product='prepro chymosin'
          FT      sig_peptide 24..71
                          /product='mat_peptide'
          FT      mat_peptide 72..1166
                          /product='mat_peptide'
          FT      mat_peptide 153..1166
                          /product='pseudochymosin'
          FT      mat_peptide 198..1166
                          /product='chymosin'.
FEATURES
source 1..1311
        location/Qualifiers
BASE COUNT 309 a 397 c 338 g 267 t
ORIGIN
Query Match 76.1%; Score 892.4; DB 6; Length 1311;
Best Local Similarity 87.4%; Pred. No. 3,9e-263;
Matches 977; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy      56 TCGTGTCTGTTACTACGCGTGTGATGATCAACCGCATTCCTCTCAAAAGTAAGTCTC 115
Db      52 TCTTCGCTCTCTCCCAAGGCGCTGAGATCACAGATCCCTCTGTCAAAAGCAAGTCTC 111
Qy      116 TCGTAAGGCGCTGAAGAAACATGACTTCTAGAAACTTTTGGCAAAACAACAGTATG 175
Db      112 TGAGGAAGGCGCTGAAGAGACATGGGCTTCTGAGAGACTTCTGCAAAACAGAGTATG 171
Qy      176 GCATACGACAGATACCTCCGCTTGGTGAAGTCTGTAGCGTGCACCTTAACCACTACC 225
Db      172 GCATACGACAGATACCTCCGCTTGGGAGGTGGCCAGCGCTGACCAACTACCC 231
Qy      236 TTGATAGTCAATCTTTGGGAAAGATCTACCTCGGAACCCGCTCAAGAGTTCAACCGTTC 295
Db      232 TGAATAGTCAATCTTTGGGAAAGATCTACCTCGGAACCCGCTCAAGAGTTCAACCGTTC 291
Qy      296 TCTTTGATACGTGTTCTCTGACTTGTGGGTTCCCTCTATCTAAGTCAAGAGTACCGTTC 355
Db      292 TGTTTGACACTGCTCTCTGACTTGTGGGTTACCTCTATCTAAGTCAAGAGTACCGTTC 351
Qy      356 GCAGAACAACCAAGATTGATCCGAGAAAGTGTCTCACTTCCAGAACTTAAGGCAAAC 415
Db      352 GCAGAACAACCAAGATTGATCCGAGAAAGTGTCTCACTTCCAGAACTTAAGGCAAGC 411
Qy      416 CTTTGTCTATACACTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 475
Db      412 CTTTGTCTATACACTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 471

```

```

QY 476 CTGCTCCAACTTGTGGACATTCAACAGAGTAGAGCTTAGACCCCAAGAACAGAGG 535
DB 472 CTGTCTCCAACTTGTGGACATTCAACAGAGTAGAGCTTAGACCCCAAGAACAGAGG 531
QY 536 ATGCTTCACTTCACTGAGAGTTCAGATGCGATCTTGTGATGAGCATACCCATGCTCGGT 595
DB 532 ACCTCTTCACTTCACTGAGAGTTCAGATGCGATCTTGTGATGAGCATACCCATGCTCGGT 591
QY 536 CAGAGTACTCCGATACCTGCTGTTGACATGATGATGATGATGATGATGATGATGATGATGAT 655
DB 532 CAGAGTACTCCGATACCTGCTGTTGACATGATGATGATGATGATGATGATGATGATGATGAT 651
QY 656 TGTCTCGGTTTACATGAGCAGAGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 715
DB 652 TGTCTCGGTTTACATGAGCAGAGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711
QY 716 ATCCATCTTCACTACACAGAGATCTTCACTGAGTTCAGATCCTGTCAGAGAGATGAGC 775
DB 712 ACCCTCTTCACTACACAGAGATCTTCACTGAGTTCAGATCCTGTCAGAGAGATGAGC 771
QY 776 AATTACCTGTGACAGTGTGACATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 835
DB 772 AGTTACCTGTGACAGTGTGACATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 831
QY 836 AAGCTATCTTGGATACCGGATCGTCCAGCTGAGTGGACCTGACAGAGAGAGAGAGAGAGAG 895
DB 832 AGGCTATCTTGGATACCGGATCGTCCAGCTGAGTGGACCTGACAGAGAGAGAGAGAGAGAG 891
QY 896 TTACAGCAAGTATTTGAGAGCAGACAGAACAGTACAGTACAGTACAGTACAGTACAGTACAG 955
DB 892 TTACAGCAAGTATTTGAGAGCAGACAGAACAGTACAGTACAGTACAGTACAGTACAGTACAG 951
QY 956 ACCTTAGCTACATGCTTACAGTGTGTTGATGATGACAGCAGAGATGATGATGATGATGATG 1015
DB 952 ACCTTAGCTACATGCTTACAGTGTGTTGATGATGATGATGATGATGATGATGATGATGATG 1011
QY 1016 CCTCGGCTTATACAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1075
DB 1012 CCTCGGCTTATACAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1071
QY 1076 ATTCCAGCAAGTATTTGAGAGCAGACAGAACAGTACAGTACAGTACAGTACAGTACAGT 1135
DB 1072 ATTCCAGCAAGTATTTGAGAGCAGACAGAACAGTACAGTACAGTACAGTACAGTACAGT 1131
QY 1136 GGGCCCAACAACTCGTGGGCTAGCTTAAGCAATCTGA 1173
DB 1132 GGGCCCAACAACTCGTGGGCTAGCTTAAGCAATCTGA 1169

```

```

RESULT 15
LOCUS E00295 1460 bp RNA linear PAT 29-SEP-1997
DEFINITION CDNA encoding pre-prorennin A.
ACCESSION E00295
VERSION E00295.1 GI:2168583
KEYWORDS JP 1985058077-A/3.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1460)
Debuttsudo, B., Romanado, U.D., Jierarudo, R.F., Arison, T.R.,
Robato, J.N., Jien, I.M., Donarudo, T.M. and Kurisuofuua, G.G.
USE OF GAL YEAST PROMOTOR
Patent: JP 1985058077-A 3 04-APR-1985;
KORABORATEIBU RES INC
OS Bos
PN JP 1985058077-A/3
PD 04-APR-1985
PF 28-FEB-1984 JP 1984035472
PR 28-FEB-1983 US 83 470911

```

```

PI DEBUTTSUDO BOTSUTOSUTEIN, RONARUDO URIN DEIBISU, PI
JIERARUDO RARUFU FUNKU,
PI ARISON TAUMTON RIGIBIT, ROBATO JIENTORII NOURUNON, JIEN I
MAO, PI DONARUDO TEIRAA MOA, KRISTOFUUA GOTSUDOFURII GORU PC
C12N15/00, C07H21/04, C07K13/00, C12N1/16, C12P21/02, C12N1/16, PC
C12R1:865,
PC (C12P21/02, C12R1:865);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue type=Stomach mucosa;
CC *source: clone=293-207 & 293-118/37;
FH Key Location/Qualifiers
FH CDS 205..1350
FT /product=pre-prorennin A'
FT mat_peptide 205..1347
FT /product=pre-prorennin A'
FT mat_peptide 253..1347
FT /product=pre-prorennin A'
FEATURES
source location/Qualifiers
1..1460
/organism="Bos taurus"
/mol_type="genomic RNA"
/db_xref="taxon:9913"
BASE COUNT 327 a 437 c 400 g 296 t
ORIGIN
Query Match 76.1%; Score 892.4; DB 6; Length 1460;
Best Local Similarity 87.4%; Pred. No. 4e-263;
Matches 977; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
56 TCGTGTCTGTATCTACCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATG 115
DB 223 TCTTGCTCTCTCCAGAGGCGCTGAGATGATGATGATGATGATGATGATGATGATGATG 292
QY 116 TCCGTAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175
DB 293 TGAGGAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
QY 176 GCATCAGCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 235
DB 353 GCATCAGCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 412
QY 236 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 295
DB 413 TGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 472
QY 296 TCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 355
DB 473 TGTTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 532
QY 356 GCAAGAACCAACCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 415
DB 533 GCAAGAACCAACCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 592
QY 416 CTTTGTCTATACACTACAGGATACAGGATACAGGATACAGGATACAGGATACAGGATACAGG 475
DB 593 CCGTGTCTATACACTACAGGATACAGGATACAGGATACAGGATACAGGATACAGGATACAGG 652
QY 476 CTGTCTCCAACTTGTGGACATTCAACAGAGTAGAGCTTAGACCCCAAGAACAGAGG 535
DB 653 CTGTCTCCAACTTGTGGACATTCAACAGAGTAGAGCTTAGACCCCAAGAACAGAGG 712
QY 536 ATGCTTCACTTCACTGAGAGTTCAGATGCGATCTTGTGATGAGCATACCCATGCTCGGT 595
DB 713 AGCTTTCACTTCACTGAGAGTTCAGATGCGATCTTGTGATGAGCATACCCATGCTCGGT 772
QY 596 CAGAGTACTCCGATACCTGCTGTTGACATGATGATGATGATGATGATGATGATGATGATGAT 655
DB 773 CAGAGTACTCCGATACCTGCTGTTGACATGATGATGATGATGATGATGATGATGATGATGAT 832

```

```
QY      656 TGTTCCTGGTTTACATGACAGGAATGGCCAGGAGACATGCTCAGCCTTGAGCTATTG 715
      |||
Db      833 TGTTCCTGGTTTACATGACAGGAATGGCCAGGAGACATGCTCAGCCTTGAGCTATTG 892
      |||
QY      716 ATTCATCTCTACTACACAGGATCTCTTCACTGGGTTCCAGTCACTGTGCAGCAGTACTGGC 775
      |||
Db      893 ACCCGTCTACTACACAGGATCTCTTCACTGGGTTCCAGTCACTGTGCAGCAGTACTGGC 952
      |||
QY      776 AATTCACTGTGACAGTGTACCATCAAGCGGTGTGTGTTGTCATGTGAAGTGTGATGTC 835
      |||
Db      953 AGTTCACTGTGACAGTGTACCATCAAGCGGTGTGTGTTGTCATGTGAAGTGTGATGTC 1012
      |||
QY      836 AAGCTTATCTTGGATACCGGTACGTTCCAGGTTGGATCGGACCTAGCAGGACATTCCTCAACA 895
      |||
Db      1013 AAGCCATCTCTGGACAGGACACCTCCAGCTGTGGGCTCCAGCAGGACATTCCTCAACA 1072
      |||
QY      896 TTCAGCAAGCTATTGAGCCACAAGAACCAAGTACGGTGAATTGACATAGATTGCGACA 955
      |||
Db      1073 TCCAGCAGGCTATTGAGCCACAAGAACCAAGTACGGTGAATTGACATAGATTGCGACA 1132
      |||
QY      956 ACCTTAGCTACATGCTCTACAGTTGTCTTTGAGATCAACGGCAAGATGTACCACTGACCC 1015
      |||
Db      1133 ACCTGAGCTACATGCTCTACAGTTGTGTGAGATCAATGCAAAATGTACCACTGACCC 1192
      |||
QY      1016 CCTCCGCTATACCAAGCAGGATCAAGGTTCTGCACCAAGTGAATCCAGAGTGAAGACC 1075
      |||
Db      1193 CCTCCGCTATACCAAGCAGGATCAAGGTTCTGTACCAAGTGAATCCAGAGTGAAGACC 1252
      |||
QY      1076 ATTCAGAAATGATCTTGGAGATGTGTCAATTCGTGAGTACTACAGCGTCTTTGACA 1135
      |||
Db      1253 ATTCAGAAATGATCTTGGAGATGTGTCAATTCGTGAGTACTACAGCGTCTTTGACA 1312
      |||
QY      1136 GGGCCAAACAACCTCGTTGGGCTAGCTAAAGCAATCTGA 1173
      |||
Db      1313 GGGCCAAACAACCTCGTTGGGCTAGCCAAAGCAATCTGA 1350
      |||
```

Search completed: February 5, 2004, 04:30:01
Job time : 4602 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 01:27:00 ; Search time 390 Seconds

(without alignments)
8119.079 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173
Sequence: 1 atgaccttccttaagctctt.....ggctagctaaagcaatctga 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1173	100.0	1173	22 AAS00569
2	1173	100.0	3957	22 AAS00570
3	898.8	76.6	1291	10 AAN91157
4	897.2	76.5	1175	5 AAN40295
5	897.2	76.5	1240	16 AAT03006
6	897.2	76.5	1460	5 AAN40180
7	895.6	76.4	1278	5 AAN40055
8	895.6	76.4	1460	3 AAN20043

9	894.8	76.3	2733	20 AA206463
10	894.8	76.3	2733	24 ABS53073
11	894	76.2	1210	12 AAQ14051
12	894	76.2	1290	4 AAN30209
13	893	76.1	2727	10 AAN91188
14	890.8	75.9	1314	4 AAN30049
15	889.8	75.9	1289	4 AAN30022
16	886.8	75.6	1098	11 AAQ04683
17	885.2	75.5	1098	4 AAN30063
18	872.8	74.4	2982	10 AAN91185
19	866.6	73.9	1173	13 AAQ20949
20	716.8	61.1	1146	24 ABS53735
21	704.6	60.1	1143	5 AAN40214
22	686.8	58.6	1104	9 AAN80001
23	534	45.5	1140	24 AAS97159
24	534	45.5	1238	24 ABS58369
25	382.4	32.6	637	14 AAQ49459
26	359.2	30.6	1393	22 AAH57421
27	359.2	30.6	1409	22 AAS44608
28	315.4	26.9	1017	23 AAS79579
29	300.8	25.6	1210	24 AAD37045
30	290	24.7	1360	20 AA220193
31	253	21.6	2158	25 ACC50980
32	253	21.6	2158	25 ABX76363
33	253	21.6	2201	24 ABZ35346
34	248.4	21.2	1357	22 AAH57422
35	248.4	21.2	1365	22 AAD09486
36	248.4	21.2	1366	24 ABK64746
37	248.4	21.2	1366	24 ABL69575
38	248.4	21.2	1418	20 AA271523
39	247.6	21.1	1340	20 AA220179
40	235.2	20.1	1285	20 AA220181
41	233.6	19.9	1258	20 AA220162
42	231.6	19.7	1130	20 AA220182
43	230.4	19.6	1173	20 AA220183
44	226.8	19.3	1173	24 AAD30576
45	226.8	19.3	1173	24 AAS97161

ALIGNMENTS

RESULT 1			
AA000569	standard: DNA, 1173 BP.		
XX	AA000569;		
XX	14-MAY-2001 (first entry)		
DT	14-MAY-2001 (first entry)		
XX	Bovine pre-pro-chymosin DNA sequence.		
DE	Bovine pre-pro-chymosin DNA sequence.		
XX	Chymosin; transcription regulator; terminator sequence; soybean; corn;		
KW	pre-pro-chymosin; ripe seed; sunflower; cotton; tobacco; alfalfa; wheat;		
KW	barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;		
KW	safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;		
KW	squash; jojoba; ds.		
OS	Bos sp.		
XX			
XX			
PH	Key	Location/Qualifiers	
FT	CDS	1..1173	
FT		/*tag= a	"Bovine chymosin"
FT		/product=	"Bovine chymosin"
FT	sig_peptide	1..78	
FT		/*tag= b	
FT	misc_feature	79..201	
FT		/*tag= c	
FT		/note=	"Pro sequence"
FT	mat_peptide	202..1170	
FT		/*tag= d	
FT		/product=	"Mature bovine chymosin"

2.7 kbp HindIII fr
DNA encoding oleos
Remain gene. Synt
Sequence of prepro
BamHI/SalI insert
cDNA sequence corr
Sequence of veal c
Sequence encoding
Sequence of proren
BamHI insert from
Prochymosin (prore
Aspergillus gene e
Gene encoding (pre
Optimised prochymo
Human aspartyl pro
protein modifiacti
Prochymosin gene 5
Human stomach cell
Human full-length
DNA encoding novel
Mouse pepsinogen-f
Cat pregnancy asso
Human bladder canc
Lung cancer-associ
Human gene express
Human stomach cell
Human gastricstin c
Human benign prost
Prostate cancer re
Human ovarian tumo
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Bovine pregnancy a
Human prolase, PR
Human aspartyl pro


```

FT      /*tag= c
FT      /note= "Phaseolin terminator"
PN      MO200114571-A1.
XX      PD
XX      01-MAR-2001.
XX      PF
XX      23-AUG-2000; 2000WO-CA00975.
XX      PR
XX      23-AUG-1999; 99US-0378696.
XX      PA
XX      (SEMB-) SEMBIOSIS GENETICS INC.
XX      PI
XX      Van Rooijen G, Keon RG, Boothe J, Shen Y;
XX      MPI; 2001-226621/23.
XX      DR
XX      P-PSDB; AAU00536.
XX      PT
XX      Producing chymosin in seeds of plants such as rice, flax, rape seed, by
XX      transforming plant cell with a nucleic acid encoding chymosin operably
XX      linked to transcription regulator and terminator sequences -
XX      PS
XX      Example 1; Fig 2; 56pp; English.
XX      CC
XX      The sequence represents a chimeric polynucleotide comprising a
XX      pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
XX      Chymosin can be produced in a plant seed through introduction of a
XX      chimeric nucleic acid molecule, comprising a nucleic acid sequence
XX      encoding a chymosin polypeptide operatively linked to transcription
XX      regulator and terminator sequences, into a plant cell. The sequences are
XX      useful for producing plant seeds, in particular seeds of soybean, rape
XX      seed, sunflower, cotton, corn, tobacco, alfalfa, barley, oats,
XX      sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm,
XX      groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
XX      rice.
SQ      Sequence 3957 BP; 1263 A; 790 C; 609 G; 1295 T; 0 other;
Query Match      100.0%; Score 1173; DB 22; Length 3957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATGAACCTCCCTTAAGCTCTTCCCTTCTACAGCTTCCCTTGGTGGTCAATATCTCGTT
DB      1554 ATGAACCTCCCTTAAGCTCTTCCCTTCTACAGCTTCCCTTGGTGGTCAATATCTCGTT
QY      61 GCTGTTACTCAAGCTGCTGAGATCAACCGCATTTCTCTCAAGAGTAAGTCTCCGT
DB      1614 GCTGTTACTCAAGCTGCTGAGATCAACCGCATTTCTCTCAAGAGTAAGTCTCCGT
QY      121 AAGGGCTGAAGGAACATGAGATTTTGAAGACTTTTGGCAAGAACAGATATGGCATC
DB      1674 AAGGGCTGAAGGAACATGAGATTTTGAAGACTTTTGGCAAGAACAGATATGGCATC
QY      181 AGCAGCAAGTACTCCGGCTTGGGTAAGTTGCTAGCGTCCACTTACCAACTACCTTGAT
DB      1734 AGCAGCAAGTACTCCGGCTTGGGTAAGTTGCTAGCGTCCACTTACCAACTACCTTGAT
QY      241 AGTCAATACTTTGGGAAGATCTACCTCGAACCCTCGCTCAAGATTCACCGTTCTCTT
DB      1794 AGTCAATACTTTGGGAAGATCTACCTCGAACCCTCGCTCAAGATTCACCGTTCTCTT
QY      301 GATACCTGTTCTCTGACATCTTGGGTTCCCTTATCTACTGCAAGCAATGCTGCAAG
DB      1854 GATACCTGTTCTCTGACATCTTGGGTTCCCTTATCTACTGCAAGCAATGCTGCAAG
QY      361 AACCAACCAAGATTCGATCCGAGAAAGTCGTCACTTCCAGAACTTAGGCAAACTTGG
DB      1914 AACCAACCAAGATTCGATCCGAGAAAGTCGTCACTTCCAGAACTTAGGCAAACTTGG
QY      421 TCTATACACTACGCTACAGGTAGCATGCAAGGAATCTTAGGCTTAGATACCGTCACTGTC
DB      1974 TCTATACACTACGCTACAGGTAGCATGCAAGGAATCTTAGGCTTAGATACCGTCACTGTC

```

```

QY      481 TCACAATTGTGACATTTCAACAGACGTAGACTTATAGCAACCCAGAACAGAGTATGTC
DB      2034 TCACAATTGTGACATTTCAACAGACGTAGACTTATAGCAACCCAGAACAGAGTATGTC
QY      541 TTCACTATGCAAGATTCGATGCGATCTTGTGATATGCAATCCATCCCTCGCTCAGAG
DB      2094 TTCACTATGCAAGATTCGATGCGATCTTGTGATATGCAATCCATCCCTCGCTCAGAG
QY      601 TACTCGATACCTGTTGTTGCAACATGATGAAACCGACACTAGTATGCTCAAGACTGTTC
DB      2154 TACTCGATACCTGTTGTTGCAACATGATGAAACCGACACTAGTATGCTCAAGACTGTTC
QY      661 TCGGTTTACATGACAGAGATGCGCAGAGAGCAATGCTCAAGCTTGAAGCTATTGATCA
DB      2214 TCGGTTTACATGACAGAGATGCGCAGAGAGCAATGCTCAAGCTTGAAGCTATTGATCA
QY      721 TCTTACTACAGAGATCTTCTCACTGCGTTCCAGTCACTGTGACAGACTAGTCCGCAATTC
DB      2274 TCTTACTACAGAGATCTTCTCACTGCGTTCCAGTCACTGTGACAGACTAGTCCGCAATTC
QY      781 ACTGAGACAGTGCACATCATGACGCGTGTGTTGTGATGTAAGTGAATGCAAGCT
DB      2334 ACTGAGACAGTGCACATCATGACGCGTGTGTTGTGATGTAAGTGAATGCAAGCT
QY      841 ATCTTGATACCGGATCGTCCAGCTGCGGATCGGACCTAGACAGCAATTCGAACATTCAG
DB      2394 ATCTTGATACCGGATCGTCCAGCTGCGGATCGGACCTAGACAGCAATTCGAACATTCAG
QY      901 CAAGCTATTTGAGGACCAACAGAACCCAGTACGCTGATTTGACATATGATTTGCAACACTT
DB      2454 CAAGCTATTTGAGGACCAACAGAACCCAGTACGCTGATTTGACATATGATTTGCAACACTT
QY      961 AGCTATATGCTTACAGTGTCTTTGAGATCAACGGCAAGATGTACCCACTGACCCCTTC
DB      2514 AGCTATATGCTTACAGTGTCTTTGAGATCAACGGCAAGATGTACCCACTGACCCCTTC
QY      1021 GCGTATACACGACGAGATCAAGGGTTCTGCACAGTGTATCCAGAGTGAAGAACATTC
DB      2574 GCGTATACACGACGAGATCAAGGGTTCTGCACAGTGTATCCAGAGTGAAGAACATTC
QY      1081 CAGAAATGATCTTGGAGATGTGTTCACTTGTGATGATCAAGCGTCTTTGACAGGGCC
DB      2634 CAGAAATGATCTTGGAGATGTGTTCACTTGTGATGATCAAGCGTCTTTGACAGGGCC
QY      1141 AACCACTCGTTGGGCTTACGTAAGCAATCTGA
DB      2694 AACCACTCGTTGGGCTTACGTAAGCAATCTGA

```

RESULT 3
 AAN91157
 ID AAN91157 standard; DNA; 1291 BP.
 XX
 AC AAN91157;
 XX
 DT 10-MAR-2003 (updated)
 DT 07-JUN-1990 (first entry)
 XX
 DE Cloned sequence of (pro)chymosin.
 DE
 KW Lactic acid bacteriæ; cheese; Streptococcus cremoris SK112; chymosin;
 KW proteinase; pSk112; ss.
 OS
 OS Unidentified.
 XX
 FH Key
 FH precursor_RNA
 FT 29..1123
 FT /*tag= a
 FT /product=:prochymosin
 PN NL8701378-A.

PD 02-JAN-1989.
XX 12-JUN-1987; 87NL-0001378.
XX 12-JUN-1987; 87NL-0001378.
XX (NEZU-) NEDERL INS ZUIVELON.
XX
XX Simons AFM, De Vos WM,
XX
XX WPI; 1989-030097/04.
XX P-PSDB; AAP94144.
XX
XX DNA fragment having region specific for lactic acid bacteria -
XX is contained in plasmid in microorganism used in prodn. of
XX protein and food prodn. eg cheese.
XX
XX Disclosure; fig 2; 43pp; Dutch.
XX
XX The DNA encoding prochymosin can be cloned into a plasmid (esp. from
XX S. cremoris SK112) and used to produce large amts of the protein by
XX recombinant DNA techniques. This could overcome the shortage of
XX prochymosin due to shortage of calf stomachs and increasing cheese
XX prodn. Prochymosin is also used in prodn. of yoghurt, butter and
XX buttermilk. See also AAN91158-N91160.
XX (Updated on 10-MAR-2003 to add missing OS field.)
XX
XX Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

Query Match 76.6%; Score 898.8; DB 10; Length 1291;
Best Local Similarity 87.7%; Pred. No. 1.2e-273;

Matches 981; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 56 TCGTGTCTGTTACTACGCTGCTAGATCACCCGATTCCTCTTACAAAGTAACTTC 115
DB 9 TCTTCGCTCTCTCCAGAGGGGCTGAGATCACAGATCCCTCTGTACAAAGGCAAGTCTC 68
QY 116 TCCGTAAAGCGCTGAGAGAAACATGACTTTAGAAAGCTTTGACAAACACAGATATG 175
DB 69 TGAGGAAGGCGCTGAGAGAGACATGGGCTTGGAGGACTTCTGACAAACAGAGATATG 128
QY 176 GCATCAGCAGCAGTACTCCGGGCTTGGTAAAGTGTAGTGCCCTTACCAACTAC 235
DB 129 GCATCAGCAGCAGTACTCCGGCTTGGAGAGTGGCGAGGTGCTTGAACCAACTAC 188
QY 236 TTGATAGTCAATCTTGGAGAGATCTACCTCGAACCCGCTCAGAGATTCACCGTTC 295
DB 189 TGAATAGTCAATCTTGGAGAGATCTACCTCGAACCCGCTCAGAGATTCACCGTTC 248
QY 296 TCTTTGATCTGCTCTCTGACTTCTGGGTTCCCTTATCTACTGCAAGACATGCTT 355
DB 249 TGTTTGATCTGCTCTCTGACTTCTGGGTTCCCTTATCTACTGCAAGACATGCTT 308
QY 356 GCAAGAACCCGCAAGATGATCCGAGAAAGTGTCAACCTTCAAGAACTTAGGCAAC 415
DB 309 GCAAGAACCCGCAAGATGATCCGAGAAAGTGTCAACCTTCAAGAACTTAGGCAAGC 368
QY 416 CCTTGTCTATATCACTAGCGTACAGTACAGTACAGAACTTATAGCTATGATCCGTCA 475
DB 369 CCTTGTCTATATCACTAGCGTACAGTACAGTACAGAACTTATAGCTATGATCCGTCA 428
QY 476 CTGTCTTCAACATTTGAGACATTTCAACAGACAGTATGAGACTTAGACCCCAAGAG 535
DB 429 CTGTCTTCAACATTTGAGACATTTCAACAGACAGTATGAGACTTAGACCCCAAGAG 488
QY 536 ATGCTTCACTATAGCAGAAATTTGATGATCCCTTGTATAGGATACCATCGCTCGCT 595
DB 489 ACGCTTCACTATAGCAGAAATTTGATGATCCCTTGTATAGGATACCATCGCTCGCT 548
QY 596 CAGAGTACTGATACCTGTGTTGACAAAGATGAAACGACACTAGTATGATCAAGACT 655
DB 549 CAGAGTACTGATACCTGTGTTGACAAAGATGAAACGACACTAGTATGATCAAGACT 608

QY 656 TGTTCGCTGTTTACATGACAGAAATGGCCAGAGAGATGCTTACAGCTTGAAGCTATTG 715
DB 609 TGTTCGCTGTTTACATGACAGAAATGGCCAGAGAGATGCTTACAGCTTGAAGCTATTG 668
QY 716 ATCCATCTACTACTACAGAGATCTTCACTGAGTTTCACTGTCAGACAGTATGAGC 775
DB 669 ACCGCTCTACTACTACAGAGATCTTCACTGAGTTTCACTGTCAGACAGTATGAGC 728
QY 776 AATTCACTGTGACAGTGTACACCATCAGCGGTGTGTTGCTTGTGATGAAAGTGTATG 835
DB 729 AGTTCACTGTGACAGTGTACACCATCAGCGGTGTGTTGCTTGTGATGAAAGTGTATG 788
QY 836 AAGCTATCTTGTATACCGGTAGCTCCAGCTGTTGCGACTAGCAGCAGATTTCAACA 895
DB 789 AGGCAATCTTGTATACCGGTAGCTCCAGCTGTTGCGACTAGCAGCAGATTTCAACA 848
QY 896 TTGAGCAAGCTATTTGAGGCAACAGAACGATGAGTGTGATGATGATGAGCA 955
DB 849 TCCAGCAGGCTTATTTGAGGCAACAGAACGATGAGTGTGATGATGATGAGCA 908
QY 956 ACCTTATGATACATGCTTACAGTGTGTTGAGATCAACGCGCAAGATGTATCCACTGACC 1015
DB 909 ACCTGAGCTACATGCTTACAGTGTGTTGAGATCAACGCGCAAGATGTATCCACTGACC 968
QY 1016 CTTCCGCTTATACCAAGCAGATCAAGGTTCTGACACAGTGTATCCAGATGAGAAC 1075
DB 969 CTTCCGCTTATACCAAGCAGATCAAGGTTCTGACACAGTGTATCCAGATGAGAAC 1028
QY 1076 ATCCCAAGAAATGATCTTGGAGATGTTGTTCACTGCTGATCTACAGGCTTTTGACA 1135
DB 1029 ATTCCTCAAGAAATGATCTTGGAGATGTTGTTCACTGCTGATCTACAGGCTTTTGACA 1088
QY 1136 GGGCCAAACAACCTGTTGGGCTAGCTTAAAGCAATCTGA 1173
DB 1089 GGGCCAAACAACCTGTTGGGCTAGCTTAAAGCAATCTGA 1126

RESULT 4
AAN40295
ID AAN40295 standard; mRNA; 1175 BP.
XX

AC AAN40295;

DT 04-FEB-1992 (first entry)

DE Sequence encoding a polypeptide displaying milk clotting activity.

XX Cheese-making; recombinant protein; rennet substitute; milk clot; ss.

OS Bos taurus.

XX Key Location/Qualifiers

FT sig_peptide 21..69

FT mat_peptide 70..1166

FT /tag= b

XX BP123928-A.

XX PD 07-NOV-1984.

XX 30-MAR-1984; 84BP-0103551.

XX PR 31-MAR-1983; 83US-0480860.

XX (CODON-) CODON GENETIC ENG.

XX Cashion LM, McCaman WT, Rice CW, Sias SR;

XX WPI; 1984-277277/45.
XX DR P-PSDB; AAP40559.
XX PT Recombinant DNA coding for milk clotting polypeptide - which is


```

QY 56 TCGTTGCTGTGTACTACGCGTGCAGATGACCCGCAATTCCTCTCAAAAGTAAGTCTC 115
DB 100 TCTTGCTCTCTCCAGGCGCGTGAAGATACAGAGATCCCTCTGTACAAAGGAATCTCTC 159
QY 116 TCCGTAAGCGCGTGAAGAACTGGAATTCTAGAAAGCTTCTTGCAAAAACAAGTATG 175
DB 160 TGAGGAAGCGGTGAAGAGACATGGGCTTCTGAGAGACTTCCGCGAAGAAACAGCATATG 219
QY 176 GCATCAGCAGCAAGTATCTCCGCTTCCGCTGAGAGTGTCTAGCGTGCACATTAACCACTACC 235
DB 220 GCATCAGCAGCAAGTATCTCCGCTTCCGCTGAGAGTGTCTAGCGTGCACATTAACCACTACC 279
QY 236 TTGATGATGAATTAATTGGGAAGATCTACTCGTGAAGCCCGGCTCAAGATTCACCGCTTC 295
DB 280 TGATAGTAGTACTTGGGAAGATCTACTCGTGAAGCCCGGCTCAAGATTCACCGCTTC 339
QY 296 TCTTGATAGTACTGTTCTCTGACTTCTGGGTTCCCTCTATCTACTGCAAGAGATGCGCT 355
DB 340 TGTGTGACACTGGCTCTCTGACTTCTGGGTTCCCTCTATCTACTGCAAGAGATGCGCT 399
QY 356 GCAAGAACCAACCAAGATTGATTCGAGAAAGTGTCCACCTTCCAGAACTTAGGCAAC 415
DB 400 GCAAAACCAACCAAGCTTCCAGCCGAGAAAGTGTCCACCTTCCAGAACTTAGGCAAC 459
QY 416 CCTGTCTATACACTACGCTACAGGTAGATGCAAGAAATCTTAGGCTATGATACCGTCA 475
DB 460 CCTGTCTATACACTACGCTACAGGTAGATGCAAGAAATCTTAGGCTATGATACCGTCA 519
QY 476 CTGTCTCCAACATTTGAGCAATTCAACAGACATGAGACTTAGCACCCAAGAACAGTIG 535
DB 520 CTGTCTCCAACATTTGAGCAATTCAACAGACATGAGACTTAGCACCCAAGAACAGTIG 579
QY 536 ATGTCTTCAACCTATGAGAAATTGATGGCATCTTGATGGCATGCCATGCTGCGCT 595
DB 580 AGGTCTTCAACCTATGAGAAATTGATGGCATCTTGATGGCATGCCATGCTGCGCT 639
QY 596 CAGAGTACCTGATACCTGTGTTGACAACTGATGAACCAAGCACTAGTATGCTCAAGACT 655
DB 640 CAGAGTACCTGATACCTGTGTTGACAACTGATGAACCAAGCACTAGTATGCTCAAGACT 699
QY 656 TGTTCCTCGTTTACATGACAGGAATGGCCAGAGAGACATGCTCACGCTTGAGAGTATTG 715
DB 700 TGTTCCTCGTTTACATGACAGGAATGGCCAGAGAGACATGCTCACGCTTGAGAGTATTG 759
QY 716 ATTCATCTTAACACAGAGATCTTCTACTGGTTCAGTCACTGTGACAGACTACTGCTG 775
DB 760 ACCCGTCTTAACACAGAGATCTTCTACTGGTTCAGTCACTGTGACAGACTACTGCTG 819
QY 776 AATTCACTGTGACAGTGTACATCAAGCGGTGTGTGTTGATGATGAAGGTGATGTC 835
DB 820 AGTTCACTGTGACAGTGTACATCAAGCGGTGTGTGTTGATGATGAAGGTGATGTC 879
QY 836 AAGCTATCTTGAATACCGGTAGCGTCAAGCTGGTGGGACTTAGACAGGACATTTCTCAACA 895
DB 880 AAGCTATCTTGAATACCGGTAGCGTCAAGCTGGTGGGACTTAGACAGGACATTTCTCAACA 939
QY 896 TTGAGCAAGCTATTTGAGCCACAAGAACCAAGTACGCTGATTTGACATGATTTGGACACA 955
DB 940 TTGAGCAAGCTATTTGAGCCACAAGAACCAAGTACGCTGATTTGACATGATTTGGACACA 999
QY 956 ACCTTGAAGTACATGCTCAAGTGTCTTTGATGAATCAAGCGCAAGATATCCCACTGACC 1015
DB 1000 ACCTTGAAGTACATGCTCAAGTGTCTTTGATGAATCAAGCGCAAGATATCCCACTGACC 1059
QY 1016 CCGTCGCGCTATACAGCGCAAGATCAAGGTTCTGACACAGTGAATTCAGAGTGAAGACC 1075
DB 1060 CCGTCGCGCTATACAGCGCAAGATCAAGGTTCTGACACAGTGAATTCAGAGTGAAGACC 1119
QY 1076 ATTCCAGAAATGATCTTGGGAGATGTGTTCAATTCGTGAGTACTACAGCGCTTTTGACA 1135
DB 1120 ATTCCAGAAATGATCTTGGGAGATGTGTTCAATTCGTGAGTACTACAGCGCTTTTGACA 1179
QY 1136 GGGCCAAACAACCTGCTGGGCTGAGCTAAGCAATCTGA 1173

```

```

DB 1180 GGGCCAAACAACCTGCTGGGCTGAGCCAAAGCATCTGA 1217

RESULT 6
AAN40180
ID AAN40180 standard; cDNA; 1460 BP.
XX
AC AAN40180;
XX
AC 25-MAR-2003 (updated)
DT 25-JAN-1992 (first entry)
XX
DE Sequence of recombinant GFP4 carrying the remain coding sequence.
XX
KW Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae;
KW se.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 205..1350
FT /*tag= a
XX
PN GB2137208-A.
XX
PD 03-OCT-1984.
XX
PF 28-FEB-1984; 84GB-0005129.
XX
PR 28-FEB-1983; 8JUS-0470911.
XX
PA (COLB ) COLLABORATIVE RES INC.
XX
PI Botstein D, Davis RW, Fink GR, Tauntontig A, Knowlton RG, Mao JI;
PI Moir DT, Goff CG;
XX
XX WPI, 1984-245517/40.
DR P-SDB; AAP40218.
XX
PT DNA segment contg. GAL1 promoter linked to gene - useful for
PT direction of expression of the gene in yeast cell
XX
PS Example; Table 4, Page 21-23; 35pp; English.
XX
CC The inventors claim a DNA segment contg. GAL1 promoter linked to
CC gene - useful for direction of expression of the gene in yeast cell.
CC The recombinant material carrying a GAL1 promoter of the yeast
CC galactokinase gene may be used in expressing a desired protein, esp.
CC bovine growth hormone, interferon, prorenin or preprorenin, in the
CC yeast cell. Strains of Saccharomyces cerevisiae producing the
CC polypeptides are produced. Yeast strains deposited as ATCC 20643,
CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
CC 528, resp. are new.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ Sequence 1460 BP; 328 A; 440 C; 399 G; 293 T; 0 other;

Query Match 76.5%; Score 897.2; DB 5; Length 1460;
Best Local Similarity 87.7%; Pred. No. 4e-273; Indels 0; Gaps 0;
Matches 980; Conservative 0; Mismatches 138;
QY 56 TCGTTGCTGTGTACTACGCGTGCAGATGACCCGCAATTCCTCTCAAAAGTAAGTCTC 115
DB 233 TCTTGCTCTCTCCAGGCGCGTGAAGATACAGAGATCCCTCTGTACAAAGGAATCTCTC 292
QY 116 TCCGTAAGCGCGTGAAGAACTGGAATTCTAGAAAGCTTCTTGCAAAAACAAGTATG 175
DB 293 TGAGGAAGCGGTGAAGAGACATGGGCTTCTGAGAGACTTCCGCGAAGAAACAGCATATG 352
QY 176 GCATCAGCAGCAAGTATCTCCGCTTCCGCTGAGAGTGTCTAGCGTGCACATTAACCACTACC 235

```

353 GCATCAGACGAGATCTCCGGCTTCGGGAGGTGCGCAGCTGCCCTGACCAATCTAC 412
 236 TTGATAGTCAATCTTTGGGAGATCTACTCGAAGCCCGCTGAGAGTTCACTGCTTC 295
 413 TGGATAGTCAATCTTTGGGAGATCTACTCGAAGCCCGCTGAGAGTTCACTGCTTC 472
 296 TCTTTGATAGTCTGCTTCGAGATCTTTGGGATCTTTGAGATCTTTGAGATCTTTG 355
 473 TGTTCGACATCTGCTTCCTGAGATCTTTGGGATCTTTGAGATCTTTGAGATCTTTG 532
 356 GCAAGAACCAACCAAGATTTCCATCCGAGAAAGTCTCCACTTCCAGAACTTGGCAAC 415
 533 GCAAAACACACAGGCTTCGACCGAGAAAGTCTCCACTTCCAGAACTTGGCAAC 592
 416 CCTTGTCTATCACTACGGTACAGGTAGATGACAGAACTTCTAGGATGATGATCCGTC 475
 593 CCTTGTCTATCACTACGGTACAGGTAGATGACAGAACTTCTAGGATGATGATCCGTC 652
 476 CTGCTCCAACTTTGAGATCTTCAACAGACAGTACGATTTAGCAACCAAGAACGGTGC 535
 653 CTGCTCCAACTTTGAGATCTTCAACAGACAGTACGATTTAGCAACCAAGAACGGTGC 712
 536 ATGTTCTTCACTATCAGATTCGATGCGATCTTTGATGCGATACCCATGCTGCGCT 595
 713 ACGTCTTCACTATCAGATTCGATGCGATCTTTGATGCGATACCCATGCTGCGCT 772
 596 CAGAGTACTCGATACCTGTTTGAACAATGATACCGAAGCTTACTAGTCTCAAGCT 655
 773 CAGAGTACTCGATACCTGTTTGAACAATGATACCGAAGCTTACTAGTCTCAAGCT 832
 656 TGTTCCTCGTTTACATGACAGAAATGCGCAGAGAGAGATGCTCAGCTTGGAGCTATTG 715
 833 TGTTCCTCGTTTACATGACAGAAATGCGCAGAGAGAGATGCTCAGCTTGGAGCTATTG 892
 716 ATTCATCTTACTACACAGAGATCTTCACTGAGTTCAGTCACTGTGACAGATCTGAC 775
 893 ACCCTCTTACTACACAGAGATCTTCACTGAGTTCAGTCACTGTGACAGATCTGAC 952
 776 AATTTCATCTGACAGTGTCTACATCAGCGCTGCTGCTTGTTCATGATGAGTGTGATG 835
 953 AGTTTACTGTGACAGTGTCTACATCAGCGCTGCTGCTTGTTCATGATGAGTGTGATG 1012
 836 AAGTATCTTGGATACCGGTACGTCAGAGTGTGCGACCTTACAGCGACATTTTCAACA 895
 1013 AGGCAATCTTGGATACCGGTACGTCAGAGTGTGCGACCTTACAGCGACATTTTCAACA 1072
 896 TTCAGCAAGCTATTGAGCCACACAGAACCACTACGCTGATGATTTGACATGATTCGACA 955
 1073 TCCAGCAAGCCATTTGAGCCACACAGAACCACTACGCTGATGATTTGACATGATTCGACA 1132
 956 ACCTTAGCTACATGCTTACAGTGTCTTGTGAGATCAACGCGCAAGATGTTACCACTGACCC 1015
 1133 ACCTGAGCTACATGCTTACAGTGTCTTGTGAGATCAACGCGCAAGATGTTACCACTGACCC 1192
 1016 CCTCGCCTATATCAGCCAGATGATCAAGGCTTCTGACCAAGTGGATTTCCAGAGTGAAGCC 1075
 1193 CCTCGCCTATATCAGCCAGATGATCAAGGCTTCTGACCAAGTGGATTTCCAGAGTGAAGCC 1252
 1076 ATTCCAGAAATGATCTTTGGAGATGTTGATCTTGTGAGTACTTACAGCGCTCTTTGACA 1135
 1253 ATTCCAGAAATGATCTTTGGAGATGTTGATCTTGTGAGTACTTACAGCGCTCTTTGACA 1312
 1136 GGGCCCAACACTCTGTTGGGCTAGTAAAGCAATCTGA 1173
 1313 GGGCCCAACACTCTGTTGGGCTAGTAAAGCAATCTGA 1350

DT 02-FEB-1992 (first entry)
 XX
 DE Sequence of prochymosin gene.
 XX
 KM Prochymosin expression vector; E.coli trp operon; chymosin; ss.
 XX
 OS Bos taurus.
 XX
 FH Key Location/Qualifiers
 XX CDS 3..1130
 FT /*tag= a
 FT 1245..1250
 FT /*tag= b
 XX
 PN EP121775-A.
 XX
 PD 17-OCT-1984.
 XX
 PF 07-MAR-1984; 84EP-0102451.
 XX
 PR 09-MAR-1983; 83JP-0038439.
 XX
 PA (BEPP/) BEPPU T.
 XX
 PI Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;
 XX Hidak M;
 DR WPI; 1984-258001/42.
 DR P-PSDB; AAP40078.
 XX
 PT Expression plasmid comprising prochymosin gene and vector -
 PT useful for transforming Escherichia coli for prochymosin prodn.
 XX
 PS Disclosure; Fig 1; 59pp; English.
 XX
 CC The inventors claim the prochymosin gene comprising a nucleotide
 CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
 CC (b) the 5th codon (GGG) to the 365th codon (ATC); and recombinant
 CC plasmids harbored by Escherichia coli strains deposited as FERM BP-
 CC 262, -263 and -264. Any portion of the nucleotide sequence as
 CC described in AAN40055 can be used. Also claimed is a vector derived
 CC from plasmid pBR322. Typically plasmid pCR501 is obt'd. from pOCT 2.
 CC The transcriptional direction of pOCT 3 is opposite to that of
 CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
 CC clockwise in pOCT 3.
 CC
 SO Sequence 1278 BP; 309 A; 384 C; 326 G; 259 T; 0 other;
 XX
 Query Match 76.4%; Score 895.6; DB 5; Length 1278;
 Best Local Similarity 87.6%; Pred. No. 1.2e-272;
 Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
 OY 56 TCGTGTGCTGTACTACAGCTGCTGAGATCACCCGATCTCTCTACAAAGTAACTCTC 115
 DB 13 TCTTCGCTCTCTCCAGGCGCTGAGATCACAGATCTCTCTGACAAAGGAGTCTC 72
 OY 116 TCCGTAAAGCCCTGAAGGATGAGTACTTCTGAAAGTCTTTGCAAAACACATATG 175
 DB 73 TGAGGAAGCCCTGAAGGATGAGTACTTCTGAAAGTCTTTGCAAAACACATATG 132
 OY 176 GCATCAGACGAGATCTCCGGCTTCGGTGAAGTTCGTAAGGTGCACTTACCAACTACC 235
 DB 133 GCATCAGACGAGATCTCCGGCTTCGGGAGGTGCGAGGTGCTTACCAACTACC 192
 OY 236 TTGATAGTCAATCTTTGGGAGATCTACTCGAAGCCCGCTGAGAGTTCACTGCTTC 295
 DB 193 TGGATAGTCAATCTTTGGGAGATCTACTCGAAGCCCGCTGAGAGTTCACTGCTTC 252
 OY 296 TCTTTGATAGTCTGCTTCGAGATCTTTGGGATCTTTGAGATCTTTGAGATCTTTG 355
 DB 253 TGTTCGACATCTGCTTCCTGAGATCTTTGGGATCTTTGAGATCTTTGAGATCTTTG 312
 OY 356 GCAAGAACCAACCAAGATTTCCATCCGAGAAAGTCTCCACTTCCAGAACTTGGCAAC 415

RESULT 7
 AAN40055
 ID AAN40055 standard; DNA; 1278 BP.
 AC
 XX AAN40055;
 XX

```

Db      313 GCAAAAACCAACGCGCTTTCACCCGAGAAAGTGTCCACCTTCCAGAACTTGGGAAGC 372
Qy      416 CCTTGTCTATACACTACGCTACAGATAGCATGCAAGGAATCTTAGCTATGATACCGTCA 475
Db      373 CCTGTCTATACACTACGCTACAGATAGCATGCAAGGAATCTTAGCTATGATACCGTCA 432
Qy      476 CTGTCTTCCAACTATTTGGACATTTCAACAGACAGTAGAGACTTAGACCCCAAGAACAGGTG 535
Db      433 CTGTCTTCCAACTATTTGGACATTTCAACAGACAGTAGAGACTTAGACCCCAAGAACAGGTG 492
Qy      536 ATGTCTTCACTATGCAAGAAATGATGAGATCTCTGGTATGAGATACCAATGCTGCGGT 595
Db      493 AGCTTTCACTATGCAAGAAATGATGAGATCTCTGGTATGAGATACCAATGCTGCGGT 552
Qy      596 CAGAGTACTGATACCTGTGTTTGAACAATGTAACCAACCACTAGTACCTAGTCAAGACT 655
Db      553 CAGAGTACTGATACCTGTGTTTGAACAATGTAACCAACCACTAGTACCTAGTCAAGACT 612
Qy      656 TGTTCCTGGTTTACATGAGACAGAAATGGCCAGAGACAGCTTCAAGCTTGAAGCTATTG 715
Db      613 TGTTCCTGGTTTACATGAGACAGAAATGGCCAGAGACAGCTTCAAGCTTGAAGCTATTG 672
Qy      716 ATCCATCTCTACTACAGAGATCTCTGCTGAGTTCAGTCACTGTGACAGATGATGAC 775
Db      673 ACCGCTCTACTACAGAGATCTCTGCTGAGTTCAGTCACTGTGACAGATGATGAC 732
Qy      776 AATTCATCTGAGACAGTGTACCATCAACGCGGTGTGTTGTGATGTAAGTGAATGTC 835
Db      733 AGTTCACTGTGAGACAGTGTACCATCAACGCGGTGTGTTGTGATGTAAGTGAATGTC 792
Qy      836 AAGCTATCTTGGATACCGGTACCTCCAACTGCTGTGCGACCTAGACAGACATTTCTCAAC 895
Db      793 AGGCAATCTTGGATACCGGTACCTCCAACTGCTGTGCGACCTAGACAGACATTTCTCAAC 852
Qy      896 TTGAGCAAGCTATTTGAGCCACACAGAACAGTACGCTGATGATTGATGATGAGCA 955
Db      853 TCCAGCAAGCTATTTGAGCCACACAGAACAGTACGCTGATGATTGATGATGAGCA 912
Qy      956 ACCTTGAATCATGCTTCAAGTGTCTTTGATGATCAACGCGCAAGATGATCCACTGACCC 1015
Db      913 ACCTGAGCTCATGCTTCAAGTGTCTTTGATGATCAACGCGCAAGATGATCCACTGACCC 972
Qy      1016 CTTCCGCTTATACAGCCAGATCAAGGGTTCTGCAACCAAGTGTCAAGTGAAGAAC 1075
Db      973 CTTCCGCTTATACAGCCAGATCAAGGGTTCTGCAACCAAGTGTCAAGTGAAGAAC 1032
Qy      1076 ATTCCAGAAATGATCTTGGGAGATGTTGTTCAATTCGTGAGTACTAGAGGTCTTTGACA 1135
Db      1033 ATTCCAGAAATGATCTTGGGAGATGTTGTTCAATTCGTGAGTACTAGAGGTCTTTGACA 1092
Qy      1136 GGGCCAAACAACCTGTGGGCTTGAAGTAAAGCAATCTGA 1173
Db      1093 GGGCCAAACAACCTGTGGGCTTGAAGTAAAGCAATCTGA 1130

```

RESULT 8
 ID AAN20043 standard; DNA; 1460 BP.
 XX AAN20043;

XX 16-DEC-1992 (first entry)

DE Pre-prorennin-A gene DNA sequence.

XX Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
 KM protease; milk-clotting enzyme; ss.

XX Bos taurus.

XX Key Location/Qualifiers
 FT CDS 205..1350

```

FT      /tag= a
FT      /label=
FT      /note= "pre-prorennin-A gene"
Db      GB2091271-A.
XX      28-JUL-1982.
Pp      15-JAN-1982; 82GB-0001120.
XX      01-DEC-1981; 81US-0325481.
PR      16-JAN-1981; 81US-0225717.
XX      (COLB ) COLLABORATIVE RES INC.
XX      Alford BL, Mao J, Moir DT;
XX      WPI; 1982-62028E/30 (62028E).
DR      P-PSDB; AAP20038.
PT      Transformed cells producing rennin and its precursors - contg.
XX      appropriate recombinant DNA material
XX      Disclosure; Table 1; 39pp; English.
CC      Bases 1-204 and 1351 to 1460 are attached to the pre-prorennin but
CC      can be removed and are not essential to use of the gene in
CC      expression. The gene may be ligated into plasmid pCGE21 and
CC      expressed in E. coli. The resulting expressed enzyme is a well
XX      known milk-clotting enzyme used in cheese-making.
SQ      Sequence 1460 BP; 329 A; 440 C; 397 G; 294 T; 0 other;
Query Match 76.4%; Score 895.6; DB 3; Length 1460;
Best Local Similarity 87.6%; Pred. No. 1.3e-272;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
Qy      56 TCGTTGCTGTACTACGCTGCTGAGATCAACCGCATTTCTCTTCAAAAGTAAGTCTC 115
Db      233 TCTTCCCTCTCTCCAGGCGCTGAGATCAACGAGATCCCTCTGTCAAAAGCAAGTCTC 292
Qy      116 TCGTAAGGGCGTGAAGAAATGACCTTCTGAAGACTTTTGGCAAAACAAGTATG 175
Db      293 TGAAGAGGGCGTGAAGAGACATGGCTTCTGAGAGACTTCTGCAAAACAGAGTATG 352
Qy      176 GCATCAGACAGAGTACTCCGCTTGTGTAAGTGTCTAGCGTCCACTTACCACTACC 235
Db      353 GCATCAGACAGAGTACTCCGCTTGTGTAAGTGTCTAGCGTCCACTTACCACTACC 412
Qy      236 TTGATAGTCANATCTTTGGAGATCTTACCTGGAAACCCGCTCAAGAGTTCAACGTTT 295
Db      413 TGAATAGTCAGTCTTTGGAGAGATCTTACCTGGAGACCCGCTCAAGAGTTCAACGTTT 472
Qy      296 TCTTATATCTGTTCTCTGACTTCTGGGTTCCCTTATCTATCTACAGCAAGCAATGCTC 355
Db      473 TGTTCACATGCTCTCTGACTTCTGGGTTCCCTTATCTATCTACAGCAAGCAATGCTC 532
Qy      356 GCAAGAACCAACAAGATTCGATCCGAGAAAGTGTGCACTTCCAGAACTTAGGCAAAAC 415
Db      533 GCAAAAACCAACGCGCTTGCACCGAGAAAGTGTGCACTTCCAGAACTTAGGCAAGC 592
Qy      416 CTTTGTCTATACACTACGCTACAGATAGCATGCAAGGAATCTTAGCTATGATACCGTCA 475
Db      593 CCTGTCTATACACTACGCTACAGATAGCATGCAAGGAATCTTAGCTATGATACCGTCA 652
Qy      476 CTGTCTTCCAACTATTTGGACATTTCAACAGACAGTAGAGACTTAGACCCCAAGAACAGGTG 535
Db      653 CTGTCTTCCAACTATTTGGACATTTCAACAGACAGTAGAGACTTAGACCCCAAGAACAGGTG 712
Qy      536 ATGTCTTCACTATGCAAGAAATGATGAGATCTCTGGTATGAGATACCAATGCTGCGGT 595
Db      713 AGCTTTCACTATGCAAGAAATGATGAGATCTCTGGTATGAGATACCAATGCTGCGGT 772

```


Db 1870 GACTTCTGGGTACCTCTATCTACTGCAAGCAAAATGCTGCAAAAACCAACGAGGCTTC 1929
 Qy 376 GATCGAGAAAGTCTCCACCTTCAGAACTTAGGCAAAACCTTGCTATACACTAGCT 435
 Db 1930 GACCCGAGAAATCGTCCACCTTCAGAACTTAGGCAAAACCTTGCTATACACTAGCT 1989
 Qy 436 ACAGGTAGCATCAAGAAATCTTAAAGCTATGATACCGTCACTGTCTCCAACTTGAGAC 495
 Db 1990 ACAGGACAGATCAAGGAGCTCTGAGGCTATGACACCGTCACTGTCTCCAACTTGAGAC 2049
 Qy 496 ATTCAACAGAGTGAAGTATGACACCAAGAAACCAAGGATGATGTCTTCACTTACAGAA 555
 Db 2050 ATCCAGCAGACAGTGAAGCTCTGAGACCCAGGAGCCCGGAGAGTCTTCACTTATCCGAA 2109
 Qy 556 TTCGATGGCATCTCTGATGAGCATACCATCGCTGCGCTGAGAGTACGATACCTG 615
 Db 2110 TTCGACGGATCTCTGAGGATGAGCTACCTCGCTGCTGCTGAGATCTGATACCTG 2169
 Qy 616 TTTGACAACTATGAAACCGACACCTAGTATGATCTCAAGACTTCTGCTGTTTACATGAC 675
 Db 2170 TTTGACAACTATGAAACCGACACCTAGTATGATCTCAAGACTTCTGCTGTTTACATGAC 2229
 Qy 676 AGGAATGCGCAGAGAGATGCTCAGCGTTGAGAGCTATTGATCATCTTACTACAGAGA 735
 Db 2230 AGGAATGCGCAGAGAGATGCTCAGCGTTGAGAGCTATTGATCATCTTACTACAGAGA 2289
 Qy 736 TCTCTTCACTGGGTTCCAGTCACTGAGCAGAGTACTGAGCAATTCATCTGAGCAGTGC 795
 Db 2290 TCTCTTCACTGGGTTCCAGTCACTGAGCAGAGTACTGAGCAATTCATCTGAGCAGTGC 2349
 Qy 796 ACCATCAGCGGTGAGTGTGTTGATGATGAGTGAAGTGTGATGATCTTGTGATACCGGT 855
 Db 2350 ACCATCAGCGGTGAGTGTGTTGATGATGAGTGAAGTGTGATGATCTTGTGATACCGGT 2409
 Qy 856 AGCTCCAGCTGCTGAGCAGTGAAGCAATCTTCAACATTCAGCAAGCTATGAGGC 915
 Db 2410 AGCTCCAGCTGCTGAGCAGTGAAGCAATCTTCAACATTCAGCAAGCTATGAGGC 2469
 Qy 916 AACAGAAACCACTAGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 975
 Db 2470 AACAGAAACCACTAGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2529
 Qy 976 GTTGTCTTGAATCAACGCGCAAGATGTACCACTGACCCCTCGCTATACAGAGCAG 1035
 Db 2530 GTTGTCTTGAATCAACGCGCAAGATGTACCACTGACCCCTCGCTATACAGAGCAG 2589
 Qy 1036 GATCAAGGGTCTGCAACAGTGAATTCAGAGTGAAGCAATTCAGAAATGATCTTG 1095
 Db 2590 GATCAAGGGTCTGCAACAGTGAATTCAGAGTGAAGCAATTCAGAAATGATCTTG 2649
 Qy 1096 GGAAGATGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
 Db 2650 GGAAGATGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2709
 Qy 1156 CTAGCTAAGCAATCTGA 1173
 Db 2710 CTAGCTAAGCAATCTGA 2727
 RESULT 11
 AAQ14051
 ID AAQ14051 standard; DNA; 1210 BP.
 XX AAQ14051;
 AC
 XX 25-MAR-2003 (updated)
 DT 06-JAN-1992 (first entry)
 XX
 DE Remnin gene.
 XX
 KW Prorennin; alpha-SI-casein gene; insulin-like growth factor I; IGF-I;
 XX mammary gland; ss.

OS Synthetic.
 XX
 PN EP451823-A.
 XX
 PD 16-OCT-1991.
 XX
 PF 10-APR-1991; 91EP-0105702.
 XX
 PR 11-APR-1990; 90DE-4012526.
 XX
 PR 11-APR-1990; 90DE-4011751.
 XX
 PA (CONE) CONSORTIUM ELEKTROCHEM IND.
 XX
 PI Harlel P, Brem G;
 DR WPI, 1991-304858/42.
 XX
 PT Recombinant DNA constructs for expressing protein in milk -
 PT contg. specific mammary gland transcription control region and
 PT signal sequence, providing high yield and easy prod. recovery
 XX
 PS Disclosure; Page 21; 41pp; German.
 XX
 CC The rennin gene was isolated using the probe represented in AAQ14775.
 CC It was used as heterologous peptide/protein together with parts of the
 CC alpha-SI-casein gene in the prodn. of DNA constructs. The heterologous
 CC peptide or protein may also be human insulin-like growth factor I.
 CC The constructs provide high yields of the protein with simple recovery
 CC from the milk. Activation of the gene occurs only in the mammary gland.
 CC See also AAQ14050, AAQ14774-77.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX
 SQ Sequence 1210 BP; 278 A; 356 G; 324 C; 252 T; 0 other;
 Query Match 76.2%; Score 894; DB 12; Length 1210;
 Best Local Similarity 87.5%; Pred. No. 3.7e-272;
 Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
 Qy 56 TCGTGTCTGTTACTACGCTGTGATGATCACCAGATCTCTCTACAAAGGTAGTCTC 115
 Db 43 TTGTGGCTGTGCTCTTGCACCTGATGATCAGAAATCCTCTGTACAAAGGCAAGTCTC 102
 Qy 116 TCCGTAAAGCGGTGAAGAAACATGACTTCTAGAACTTCTTGCAAAACAGATATG 175
 Db 103 TGAGGAAGCGGTGAAGAAAGATGAGCTTCTGAGAGACTTCTTGCAAAACAGATATG 162
 Qy 176 GCATCAGCAGCAAGTACTCCGGCTTGGTGAAGTGTCTAGCGTGCACCTTACCAATACC 235
 Db 163 GCATCAGCAGCAAGTACTCCGGCTTGGTGAAGTGTCTAGCGTGCACCTTACCAATACC 222
 Qy 226 TTGATATGATCAATCTTGGGAAAGATCTTACCTCGAAACCCCGCTCAAGATTCACGTTT 295
 Db 223 TGGATATGATCAATCTTGGGAAAGATCTTACCTCGAAACCCCGCTCAAGATTCACGTTT 282
 Qy 296 TCTTGTATATGAGTCTCTGATCTTCTGAGGTTCCCTTATCTACTGCAAGAGCAATGCTT 355
 Db 283 TCTTGTATATGAGTCTCTGATCTTCTGAGGTTCCCTTATCTACTGCAAGAGCAATGCTT 342
 Qy 356 GCAGAAACCAACCAAGATTCGAGTCCGAAAGTGTCTGACCTTCCAGAACTTAGGCAAA 415
 Db 343 GCAGAAACCAACCAAGATTCGAGTCCGAAAGTGTCTGACCTTCCAGAACTTAGGCAAA 402
 Qy 416 CTTGTCTATATGATCAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 475
 Db 403 CTTGTCTATATGATCAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 462
 Qy 476 CTGTCTCAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 535
 Db 463 CTGTCTCAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 522
 Qy 536 ATGTCTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 595
 Db 523 ATGTCTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582


```

QY 596 CAGAGTACTGATACCTGTGTGTTGACAACTGATGAACCGACACTAGTACCTCAGACT 655
DB 583 CAGAGTACTGATACCTGTGTGTTGACAACTGATGAACCGACACTGTGTGCCAAGACC 642
QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGACATGCTCAGCGCTTGAGCTATTG 715
DB 643 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGACATGCTCAGCGCTTGAGCGCATCG 702
QY 716 ATCCATCTCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGACTAGTGGC 775
DB 703 ACCGCTCTCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGACTAGTGGC 762
QY 776 AATTCACTGTGACAGTGTACCATCAGCGGTGTGTGTTGATGATGGAAGTGTGATGTC 835
DB 763 AGTTCACTGTGACAGTGTACCATCAGCGGTGTGTGTTGATGATGGAAGTGTGATGTC 822
QY 836 AACCTATCTTGGATACCGGTACGTCGAAGCTGGTGGGACCTAGCAGGACATCTCAGACA 895
DB 823 AGGCACTCTGACACGAGGACCTCCAGCTGGTGGGACCGACGATATCTCTCAGACA 882
QY 896 TTCAGCAAGCTATTTGAGCCACACAGAACAGTACGCGTGAAGTTGACATGATTTGCGACA 955
DB 883 TTCAGCAAGCTATTTGAGCCACACAGAACAGTACGCGTGAAGTTGACATGATTTGCGACA 942
QY 956 ACCTTAGCTACATGCTCTACAGTGTCTTTGAGATCAACGCGAAGATGTACCCACTGACCC 1015
DB 943 ACCTGAGCTACATGCTCTACAGTGTCTTTGAGATCAACGCGAAGATGTACCCACTGACCC 1002
QY 1016 CCTCCGCTTATACAGCCAGAGATCAAGGGTTCTGACACAGTGAATTCAGAGTGAGAAC 1075
DB 1003 CCTCCGCTTATACAGCCAGAGATCAAGGGTTCTGACACAGTGAATTCAGAGTGAGAAC 1062
QY 1076 ATTCCAGAAATGATCTTGGAGATGTCTTCACTTCTGATCTACAGCGCTTTTGACA 1135
DB 1063 ATTCCAGAAATGATCTTGGAGATGTCTTCACTTCTGATCTACAGCGCTTTTGACA 1122
QY 1136 GGGCCAAACAACCTGCTGGGTAGCTAAAGCAATCTGA 1173
DB 1123 GGGCCAAACAACCTGCTGGGTAGCTAAAGCAATCTGA 1160

RESULT 12
AAN30209
ID AAN30209 standard; cDNA; 1290 BP.
AC AAN30209;
XX
DT 25-MAR-2003 (updated)
DT 03-AUG-1992 (first entry)
DE
XX Sequence of preprochymosin cDNA.
XX
XX Milk-clotting; cheese making; enzyme; zymogen; ss.
XX
XX Cow.
XX
XX Key Location/Qualifiers
XX CDS 26..73 /*tag= a
XX FT /*tag= b
XX FT /label= preprochymosin
XX CDS 74..199
XX FT /*tag= b
XX FT /label= prochymosin
XX CDS 200..1171
XX FT /*tag= c
XX FT /label= chymosin
XX
XX GB2100737-A.
XX PN
XX 06-JAN-1983.
XX PD
XX 11-JUN-1982; 82GB-0017096.

```

```

XX 11-JUN-1982; 82GB-0017096.
PR (CLUT ) CELITECH LTD.
PA
XX Carey NH, Harris TUR, Lowe PA, Doel MT, Emtage JS;
PI WPI; 1983-00545K/01.
XX DR P-PSDB; AAP30446.
XX
XX Prod. or calf stomach chymosin for cheese making - by
PT cultivation of micro-organisms transformed with vector system
XX
XX Claim 42; Fig 4; 26pp; English.
XX
XX The inventors claim a method for the prodn. of calf stomach chymosin
CC for cheese making. Genes and polypeptides for preprochymosin,
CC prochymosin and chymosin are claimed, as are vector systems and a
CC prochymosin primer.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 1290 BP; 309 A; 384 C; 334 G; 263 T; 0 other;

Query Match 76.2%; Score 894; DB 4; Length 1290;
Best Local Similarity 87.5%; Pred.No.3,8e-272;
Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 56 TCGTTCCTGTTACTACGCTGCTGAGATCACCGCATTCCTCTCTACAAAGTAACTTC 115
DB 54 TCTTCGCTCTCTCCCAAGGCGCTGAGATCACAGATTCCTCTCTCTACAAAGCACTTC 113
QY 116 TCCGTAAGGCGCTGGAAGAACTGACTCTTGAAGACTTTCTTGCAAAACAACAGTATG 175
DB 114 TGAGAAAGGCGCTGGAAGAGACTGAGGCTTCTGAGAGACTTCTTGCAAAACAAGCATATG 173
QY 176 GCATCAGCAGCAAGTACTCCGCTTGGTGAAGTGTCTAGGCGTCCAGCACTTACCAATACC 235
DB 174 GCATCAGCAGCAAGTACTCCGCTTGGGAGGTGCGACGCGCTTGACCAACTTACC 233
QY 236 TTGATAGTCAATCTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGATTCCACGGTTC 295
DB 234 TAGATAGTCAATCTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGATTCCACGGTTC 293
QY 296 TCTTGATATGCTGTTCTCTGACTTCTGGGTTCCCTCTATCTACTGCAAGAGCAATGCTC 355
DB 294 TGTTCGACTGCTCTCTGACTTCTGGGTTCCCTCTATCTACTGCAAGAGCAATGCTC 353
QY 356 GCAGAAACCAACCAAGATTCGATCCGAGAAAGTCTCACCTTCGAAACTTATAGGCAAC 415
DB 354 GCAGAAACCAACCAAGATTCGATCCGAGAAAGTCTCACCTTCGAAACTTATAGGCAAGC 413
QY 416 CCTTGTCTATACCTACGCTGAGTACGATGCAAGAAATCTTAGGCTATGATACCGTCA 475
DB 414 CCTTGTCTATACCTACGCTGAGTACGATGCAAGAAATCTTAGGCTATGATACCGTCA 473
QY 476 CTGTCTCCAACATTTGAGCAATTCACACAGACATGAGACTTTAGCACCAAGAACAGGTG 535
DB 474 CTGTCTCCAACATTTGAGCAATTCACACAGACATGAGACTTTAGCACCAAGAACAGGTG 533
QY 536 ATGTCTTCACTTATGCAAGATTTGATGAGCATCTTGTGATGATATCCATGCTGCGGT 595
DB 534 ACGTCTTCACTTATGCAAGATTTGATGAGCATCTTGTGATGATATCCATGCTGCGGT 593
QY 596 CAGAGTACTGATACCTGTGTGTTGACAACTGATGAACCGACACTAGTACCTCAGACT 655
DB 594 CAGAGTACTGATACCTGTGTGTTGACAACTGATGAACCGACACTAGTACCTCAGACT 653
QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGACATGCTCAGCGCTTGAGCTATTG 715
DB 654 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGACATGCTCAGCGCTTGAGCGCATCG 713
QY 716 ATCCATCTCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGACTAGTGGC 775

```


Db 714 ACCCTCTACTACACAGGTCCTCCGCACTGGGTCCCGTGACAGTGCAGCTACTGTC 773
Qy 776 AATTACTGTGACAGTGTGCACATCAGCGGTTGTTGTCATGTAAGGTGATGTC 835
Db 774 AGTTACTGTGACAGTGTGCACATCAGCGGTTGTTGTCATGTAAGGTGATGTC 833
Qy 836 AAGCTATCTGTGATACCGGTACGTTCAGAGTGTGTCGACTTACAGCAGCAATTCACAA 895
Db 834 AGGCATCTTGGACACGGGCACTCTCAAGCTGTGGGCCCCAGCAGCAGCACTCTCAAC 893
Qy 896 TTCAGCAAGCTATTGGAGCCACACAGAACAGTACGGTGAAGTTGACATAGATTCCACA 955
Db 894 TCAGACAGCCATTTGGAGCCACACAGAACAGTACGGTGAAGTTGACATAGATTCCACA 953
Qy 956 ACCTAGCTACATGCTACAGTGTCTTGGATCAAGGGCAAGTATACCACTGACCC 1015
Db 954 ACCTAGCTACATGCTACAGTGTCTTGGATCAAGGGCAAGTATACCACTGACCC 1013
Qy 1016 CCTCGCTATATACAGCCAGATCAAGGTTCTGACAGTGCATTCAGAGTGAAC 1075
Db 1014 CCTCGCTATATACAGCCAGATCAAGGTTCTGACAGTGCATTCAGAGTGAAC 1073
Qy 1076 ATTCCAGAAATGATCTTGGAGATGTTGATTCATTCAGTATACAGCGCTTTGACA 1135
Db 1074 ATTCCAGAAATGATCTTGGAGATGTTGATTCATTCAGTATACAGCGCTTTGACA 1133
Qy 1136 GGGCCACACACTCGTGGGCTAGCTAAAGCAATCTGA 1173
Db 1134 GGGCCACACACTCGTGGGCTAGCTAAAGCAATCTGA 1171

RESULT 13

ID AAN91188 standard; DNA; 2727 BP.

XX AAN91188;
XX
XX
XX 25-MAR-2003 (updated)
XX 15-JUN-1990 (first entry)
XX
XX BamHI/Sall insert of Kluyveromyces plasmid PAB309.
XX
XX Kluyveromyces; pBM100PC; chymosin; tissue plasminogen activator.
XX
XX Kluyveromyces lactis.
XX
XX Key Location/Qualifiers
XX CDS 409..1781
XX /*tag= a
XX
XX EP301670-A.
XX
XX 01-FEB-1989.
XX
XX 28-JUL-1988; 88BP-0201632.
XX
XX 28-JUL-1987; 87US-0078539.
XX
XX (KONN) GIST-BROCADES NV.
XX
XX Vandenberg JA, Vanooijen AJU, Rietveld K;
XX
XX WPI; 1989-033565/05.
XX P-PSDB; AAP94376.
XX
XX Kluyveromyces host cells for producing polypeptides) -
XX PT used for highly efficient prodn. of eg chymosin tissue
XX PT plasminogen activator or human serum albumin.
XX
XX Disclosure; Page ?; 56pp; English.
XX
XX BamHI/Sall insert, incorporated into plasmids PAB309 with

CC promoters, terminators and a G418 resistance marker fused to a ADHI
CC promoter from S.cerevisiae.
CC The Kluyveromyces expression systems provide highly efficient secretion
CC and processing of a wide variety of proteins.
CC Sequences identical to those published in EP301669.
CC (Updated on 25-MAR-2003 to correct PD field.)
CC
CC (Updated on 25-MAR-2003 to correct PI field.)
CC
SQ Sequence 2727 BP; 752 A; 654 C; 572 G; 749 T; 0 other;

Query Match 76.1%; Score 893; DB 10; Length 2727;
Best Local Similarity 87.5%; Pred. No. 1,2e-271;
Matches 977; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 57 CGTTCCTGTTACTACCGCTGTGAGATACCCGCAATTCCTCTCAAAAGGTAAGTCTCT 116
Db 669 CGATGCTTCCATCATATATGCTTGAGATCACCAGATCCCTTTGACAAAGGCAAGTCTCT 728
Qy 117 CCGTAGGCGCTGAAGAAACATGACCTTCTAGAACTTCTTGAAGAAACAAAGATATG 176
Db 729 GAGGAAGGCGCTGAAGAGCATGGGCTTCTGAGAGACTTCTGCAAGAAACAGAGATATG 788
Qy 177 CATGACGACAGTACTCCGGCTTCGGTGAAGTTGCTAGCGTGCCTTACCAACTACT 236
Db 789 CATGACGACAGTACTCCGGCTTCGGGAGTGGCCAGCGCTCCCTGACCAACTACT 848
Qy 237 TGATAGTCAATCTTGGGAGATCTACTCGGAACCCCGGCTCAAGAGTTCAACCGTCT 296
Db 849 GAGACGTAGTACTTGGGAGATCTACTCGGAACCCCGGCTCAAGAGTTCAACCGTCT 908
Qy 297 CTTGATACTGGTCTCTGACTTCTGAGTTCCCTCTATCTACTGCAAGACAAATGCTCT 356
Db 909 GTTGACACTGGCTCTCTGACTTCTGAGTTCCCTCTATCTACTGCAAGACAAATGCTCT 968
Qy 1029 CCTGCTATCCACTACGAGGACGAGCATGACGGGCACTCGGCTATGACACCGTCA 1088
Db 417 CTTGCTATACACTACGATACAGTATGATGATGATGATGATGATGATGATGATGATG 476
Qy 477 TGTCTCCAACTTGTGACATTCACAGACAGTATGACTTACAGCCCAAGAACAGTATG 536
Db 1089 TGTCTCCAACTTGTGACATTCACAGACAGTATGACTTACAGCCCAAGAACAGTATG 1148
Qy 537 TGTCTCCAACTTGTGACATTCACAGACAGTATGACTTACAGCCCAAGAACAGTATG 596
Db 1149 GGTCTTCACTATGCGCAATTCGACCGGATCTGGGATGAGCTTACCTCGCTGCTC 1208
Qy 597 AGAGTACTGATACCTGTGTTGACACATGATGAAACCGACACTAGTACTGAAGACTT 656
Db 1209 AGAGTACTGATACCTGTGTTGACACATGATGAAACCGACACTAGTACTGAAGACTT 1268
Qy 657 GTTCTCGGTTTAACTGACAGAGATGCGCAGAGAGCATGCTACCGTTGAGCTATGTA 716
Db 1269 GTTCTCGGTTTAACTGACAGAGATGCGCAGAGAGCATGCTACCGTTGAGCTATGTA 1328
Qy 717 TCATCTCTACTACAGAGATCTTCACTGGGTTCACTGACCTGACAGTATGCA 776
Db 1329 CCGCTCTACTACAGAGATCTTCACTGGGTTCACTGACCTGACAGTATGCA 1388
Qy 777 ATTCACTGTGACAGTGTACCATCAGCGGTGTGTTGATGATGAAAGTGTATGCA 836
Db 1389 GTTCACTGTGACAGTGTACCATCAGCGGTGTGTTGATGATGAAAGTGTATGCA 1448
Qy 837 AGCTATCTTGAATACCGGTAGCTGCAAGCTGCTGCAAGCTGCAAGCTGCAAGCT 896
Db 1449 GGCATCTGTGACAGGCACTTCAAGCTGCTGCAAGCTGCAAGCTGCAAGCT 1508
Qy 897 TCAGCAAGCTATTTGAGGACACAGAACAGTACGGTATGTTGACATAGATTGCA 956
Db 1509 CAGAGGCACTTTGAGGACACAGAACAGTACGGTATGTTGACATAGATTGCA 1568

```
QY 957 CCTAGCTACATGCTCAGATGTTGTTTGGATCAACGCAAGATGTTACCCACTGACCCC 1016
DB 1569 CTTGAGCTACATGCCCACTGTGTTCTTTGATCAATGCAAAATGTACCCACTGACCCC 1628
QY 1017 CTCGGCTATACAGCAGAGATCAAGGGTTCGCACCGATGATTTCCAGATGAGAACCA 1076
DB 1629 CTCGGCTATACAGCAGAGATCAAGGGCTTCTGTACAGAGTGGCTTCCAGATGAAATCA 1688
QY 1077 TTCCCAAGAAATGATCTTGGAGATGTGTTTCATCTCGAGTACTACAGCGCTTTGACAG 1136
DB 1689 TTCCCAAGAAATGATCTTGGGGATGTGTTTCATCTCGAGATATTACAGCGCTTTGACAG 1748
QY 1137 GGCCAAACAACCTGTGGGCTAGCTAAAGCAATCTGA 1173
DB 1749 GGCCAAACAACCTGTGGGCTAGCCCAAGCCATCTGA 1785

RESULT 14
ID AAN30049 standard; cDNA; 1314 BP.
AC AAN30049;
XX 14-JUN-1992 (first entry)
DT
XX cDNA sequence corresponding to one of the allelic forms (B) of
DE bovine preprochymosin.
XX Chymosin; enzyme; rennet; cheese; ss.
XX
XX Bos taurus.
OS
XX
XX Key Location/Qualifiers
FH sig_peptide 24..40
FT /*tag= a
FT mat_peptide 41..1169
FT /*tag= b
XX
XX BP77109-A.
XX
XX 20-APR-1983.
XX
XX 13-OCT-1982; 82BP-0201272.
XX
XX 14-OCT-1981; 81GB-0031004.
XX
XX (UNIL ) UNILEVER NV.
XX
XX PA
XX Maat J, Verrips CT, Ledebor AM, Edens L;
XX
XX WPI; 1983-39656K/17.
XX
XX P-PSDB; AAP30086.
XX
XX DNA molecules comprising genes for preprochymosin - used to
PT transform microorganisms to give strain producing the
PT prepro-enzyme and its allelic and maturation forms
XX
XX Claim 2; Fig 1; 53bp; English.
XX
XX Preprochymosin is an intermediate (via prochymosin and
CC pseudochymosin) for the enzyme chymosin, which is the essential milk-
CC clotting component of rennet and is used in cheese manufacture.
CC AAN30049 corresp. to mRNA isolated from the fourth stomach of a
CC preumant calf (abomasum, Friesian cow).
XX
XX Sequence 1314 BP; 309 A; 398 C; 338 G; 269 T; 0 other;
SO
Query Match 75.9%; Score 890.8; DB 4; Length 1314;
Best Local Similarity 87.3%; Pred. No. 4e-271;
Matches 976; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 56 TCGTGTCTTACTACAGCTGTGAGATCAACCCGCAATTCCTCTCTACAAAGTAAGTCTC 115
```

```
DB 52 TCTTGCCTCTCTCCAAAGCGCTGAGATTCACAGAGATCCCTCTGTATCAAAAGCAAGTCTC 111
QY 116 TCCGTAAAGCGCTGAAGAAACATGACCTTGTAAAGACTTTTGGAAACAAACAGTATG 175
DB 112 TGAGAAAGCGCTGAAGGAGCATGCGCTTCTGAGAGACTCTTCGAGAAACAGCAGTATG 171
QY 176 GCATCAAGCAGAAATCTCCGCGCTTCGGAAGTTGTATGCGGTGCACCTTACCAATTACC 235
DB 172 GCATCAAGCAGAAATCTCCGCGCTTCGGAAGTTGTATGCGGTGCACCTTACCAATTACC 231
QY 236 TTGATAGTCAATATCTTGGGAAGATCTACCTCGAAACCCGCGCTCAAGATTACCGTTTC 295
DB 232 TGATATGATAGTACTTGGGAAGATCTACCTGGGAGCCCGGCCAGAGATTACCGTTTC 291
QY 296 TCTTTGATATCTGTTCTCTGACTTCTGCGTTCCCTGTATCTACTGCAAGAGCAATGCTT 355
DB 292 TGTTTGACACTGCGCTCTCTGACTTCTGCGTTCCCTGTATCTACTGCAAGAGCAATGCTT 351
QY 356 GCAAGAACCAACCAAGATTCGATTCGAGAAAGTCGTCACCTTCCAGAACTTAAGGCAAC 415
DB 352 GCAAAACCAACCAAGCGCTTTCGACCGAGAAAGTCGTCACCTTCCAGAACTTGGCAAGC 411
QY 416 CTTTGTCTATACACTACGCTAGCAGTACGATGCAAGAACTTAAAGCTATGATACCGTCA 475
DB 412 CCGTGTCTATACACTACGCGAGCAGAGCATGCAAGGCAATCTGGGCTATGACACCGTCA 471
QY 476 CTGTCTCCAAACATTGTGACATTCAACAGACAGTAAAGATTAGCAACCAAGAACAGGTG 535
DB 472 CTGTCTCCAAACATTGTGACATTCAACAGACAGTAAAGCTTGAAGCCCTGAGACCCAGAGCCCGGG 531
QY 536 ATGCTTTCACCTATGAGAAATTCAGATGCAATCTTGATATGCAATGCAATGCTGCGCT 595
DB 532 AGCTTTCACCTATGAGCAATTCAGAGATTCGAGAAAGTTCGAGAAAGCTTCCCTGCTGCGCT 591
QY 596 CAGAGTACTGATACCTGTGTTGACCAATGATGAACCGACACTAGTATGCTCAAGACT 655
DB 592 CACAGTACTGATACCGGTGTTGACCAATGATGAACCGACACTGATGCTCAAGACT 651
QY 656 TGTTCGCTTACATGACAGAAATGCTCAGAGAGCATGCTCAGCTTGGAGCTATTG 715
DB 652 TGTTCGCTTACATGACAGAAATGCTCAGAGAGCATGCTCAGCTTGGAGCTATTG 711
QY 716 ATTCATCTTACTACAGATCTCTTCACTGGGTTCCATCTGTCAGCAGTATGTC 775
DB 712 ACCGTCCTTACTACAGAGGCTCTGCAATGAGTGGTCCGTCAGAGTCAAGTATGTC 771
QY 776 AATTCACTGAGCAGATCTCAACATCAGCGGTGTGTTGTTGATGATGAAGGTGATGTC 835
DB 772 AGTTCACTGAGCAGATCTCAACATCAGCGGTGTGTTGTTGATGATGAAGGTGATGTC 831
QY 836 AAGCTATCTTGAATACCGGTACGTCCTCAAGCTGTCGACCTTACAGCAGCATTTCTTACA 895
DB 832 AGGCAATCTTGAACAGGCGACCTCCAGCTGATGCGGCCAGCAGCATCTCTACA 891
QY 896 TTGAGCAAGTATGAGGCAACAGAAACAGTACGATGATTTGACATAGATTGGACA 955
DB 892 TCCAGAGGCAATGAGGCAACAGAAACAGTACGATGATTTGACATAGATTGGACA 951
QY 956 ACCTAGCTACATGCTCAGTATGTTCTTGAATCAACGGCAAGATGATCCACTGACC 1015
DB 952 ACTTAGCTACATGCTCAGTATGTTCTTGAATCAATGGCAAAATGATCCACTGACC 1011
QY 1016 CTTCCGCTATACAGCAGAGATCAAGGGTTCGACCAAGTATTCAGAGTGAAGACC 1075
DB 1012 CTTCCGCTATACAGCAGAGATCAAGGGCTTCTGTATCAAGTGTCTTCAAGTGAAGAAC 1071
QY 1076 ATTCCCAAGAAATGATCTTGGAGATGTGTTATCTGTAGTACTACAGCTTTTGA 1135
DB 1072 ATTCCCAAGAAATGATCTTGGAGATGTGTTATCTGTAGTACTACAGCTTTTGA 1131
QY 1136 GGGCCAAACAACCTGTGGGCTAGCTAAAGCAATCTGA 1173
```

Db 1132 GGGCCAAACCTCGTGGGGCTGGCCAAAGCCATCTGA 1169

RESULT 15
AAN30022
ID AAN30022 standard; DNA; 1289 BP.

AC AAN30022;
XX
XX
DT 25-MAR-2003 (updated)
DT 25-APR-1992 (first entry)
XX
XX Sequence of veal chymosin gene.
DE
XX Protolytic enzyme; zymogen; rennin; chymosin; cheese making;
KW microbial vector; ss.
XX
XX Bos taurus.
OS
XX
XX Key Location/Qualifiers
FH 1.1140
FT CDS /*tag= a
ET
XX
XX BE897201-A.
XX
XX PD 03-NOV-1983.
XX
XX PF 30-JUN-1983; 83BE-0017731.
XX
XX PR 13-APR-1983; 83US-0484539.
XX PR 01-JUL-1982; 82US-0394433.
XX
XX PA (GENEX) GENEX CORP.
XX
XX DR WPI, 1983-820813/47.
XX P-PSDB; AAP30013.
XX
XX PT Isolated chymosin or rennin and prochymosin genes - plasmid(s)
PT which replicate in prokaryotic organisms, esp. Escherichia coli,
PT and organisms used for chymosin biosynthesis
XX
XX PS Claim 7, Page 33-36; 43pp; French.
XX
XX CC The inventors claim isolated chymosin (rennin) and prochymosin genes
CC from calves, and plasmids contg. the genes which are capable of
CC replicating in a prokaryotic organism. The prokaryotic organism is
CC pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).
CC The microorganisms transformed by the plasmid are also claimed.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX SQ Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;

Query Match 75.9%; Score 889.8; DB 4; Length 1289;
Best Local Similarity 87.2%; Pred. No. 8.2e-271;
Matches 975; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 56 TCGTGTCTTACTCAAGCTGCTGAGATACCCGCACTTCTCTCAAAAGGTAAGTCTC 115
DB 23 TCTTGTCTCTCTCCAGGGCGCTGAGATCACAGGATCCCTCTGTAACAAAGGCAAGTCTC 82

QY 116 TCGTAAGCGCGTGAAGAACTGACTCTTAAGAACTTCTTGAGAAACAAAGCATGATG 175
DB 83 TGAGGAAGCGCTGAAGAGCATGGCTTCTGAGAGACTTCTGAGAAACAGCACTCTG 142

QY 176 GCATGAGCAGCAAGTACTCGGCTTCGATGAAGTTGCTAGCGTGCACCTTACCACTACC 235
DB 143 GCATGAGCAGCAAGTACTCGGCTTCGAGGACGTGGCCAGCGTGCCTGACCACTACC 202

QY 236 TTGATAGTAATTAATCTTGGGAAGATCTACTCGGAACCCCGCTCAAGATTCACCGTTC 295
DB 203 TAGATAGTCAATCTTGGGAAGATCTACTCGGAGCCCGCCCAAGGATTCACCGTTC 262

QY 296 TCTTGTACTGATGCTCTGACTTCTGAGTTCCTCTATCTACTGCAAGAGCAATGCTC 355

Db 263 TGTGTGACACTGGCTCCTCTCTGACTTCTGGGTACCTCAATCTACTGCAAGAGCAATGCTC 322

QY 356 GGAAGAACACCAAAAGATTGGATCCGAGAAAGTGTGTCACCTTCAGAACTTAGGCAAC 415

DB 323 GGAAGAACACCAAGGCTTTCAGACCCGAGAAAGTGTGTCACCTTCCAGAACTTAGGCAAGC 382

QY 416 CTTGTCTATACACTACCGTACAGGTAGCATGACAGAAATCTTAGGCTATGATACCGTCA 475

DB 383 CCTGTCTATCACTACAGGAGACAGGACATGACAGGAGCATCTGGGCTATGACCGTCA 442

QY 476 CTGTCTCAACATTTGTGGAATTTCAACAGACATGATGACTTAGCAACCAAGAACAGGTG 535

DB 443 CTGTCTCAACATTTGTGGAATTTCAACAGACATGATGAGCTTAGCAACCAAGAACAGGTG 502

QY 536 ATGCTTCACTTACGACAAATTTCAATGAGCATCTTGTATGATGACCAATCCATGCTCGGT 595

DB 503 AGCTTCACTTACGACAAATTTCAATGAGCATCTTGTATGATGACCAATCCATGCTCGGT 562

QY 596 CAGAGTACTGATACCTGTGTTTGAACATGATGATGACCACTAGTACCTAGCAAGACT 655

DB 563 CAGAGTACTGATACCTGTGTTTGAACATGATGATGACCACTAGTACCTAGCAAGACT 622

QY 656 TGTCTCGGTTTACATGAGCAGAAATGAGCAGAGAGAGATGCTACGCTTGGAGCTATTG 715

DB 623 TGTCTCGGTTTACATGAGCAGAAATGAGCAGAGAGAGATGCTACGCTTGGAGCTATTG 682

QY 716 ATTCATCTTACTACACAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGCATGTGC 775

DB 683 ACCGCTCTTACTACACAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGCATGTGC 742

QY 776 AATTCATCTTGGAGCAGTGTCAACATCAGCGGTGTTGTTGATGATGATGATGATGTC 835

DB 743 AGTTCATCTTGGAGCAGTGTCAACATCAGCGGTGTTGTTGATGATGATGATGATGTC 802

QY 836 AAGTATCTTGGATACCGGTACGTCCAGCTGTGTCGACCTAGCAGCAGCAATTTCAACA 895

DB 803 AGGCATCTTGGATACCGGTACGTCCAGCTGTGTCGACCTAGCAGCAGCAATTTCAACA 862

QY 896 TTCAGCAAGCTATTTGAGACCAACAGAACAGTACGATGATTTTACATGATTTGCGACA 955

DB 863 TTCAGCAAGCTATTTGAGACCAACAGAACAGTACGATGATTTTACATGATTTGCGACA 922

QY 956 ACCTAGCTATGATGCTTACAGTGTCTTGAATCAAGGCAAGATGATACCACTGACCC 1015

DB 923 ACCTAGCTATGATGCTTACAGTGTCTTGAATCAAGGCAAGATGATGATACCACTGACCC 982

QY 1016 CCTCGCTTATACAGCAGGATCAAGGTTCTGACCAAGTGTGATTCAGAGTGAAC 1075

DB 983 CCTCGCTTATACAGCAGGATCAAGGTTCTGACCAAGTGTGATTCAGAGTGAAC 1042

QY 1076 ATTCCGAAGATGATCTTGGAGATGTTGTTGATGATGATGATGATGATGATGATGATG 1135

DB 1043 ATTCCGAAGATGATCTTGGAGATGTTGTTGATGATGATGATGATGATGATGATGATG 1102

QY 1136 GGGCCAAACCTCGTGGGCTAGCTAAAGCAATGTA 1173

DB 1103 GGGCCAAACCTCGTGGGCTAGCTAAAGCAATGTA 1140

Search completed: February 5, 2004, 03:13:10
Job time : 398 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 02:57:45 ; Search time 92 Seconds
(without alignments)
5627.635 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173

Sequence: 1 atgactccttaagctctt.....ggctagctaaagcaatcgtga 1173

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues 1139956

Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	897.2	76.5	1240	US-08-240-372-2	Sequence 2, Appli
2	894.8	76.3	2732	US-08-846-021A-6	Sequence 6, Appli
3	882	75.2	2732	5217891-14	Patent No. 5217891
4	203.6	17.4	832	US-08-318-193-15	Sequence 15, Appli
5	203	17.3	828	5200327-6	Patent No. 5200327
6	199.8	17.0	846	5200327-5	Patent No. 5200327
7	199.4	17.0	838	US-08-318-193-13	Sequence 13, Appli
8	192.2	16.4	2073	US-09-032-523-6	Sequence 6, Appli
9	146.6	12.5	1299	US-08-723-938-4	Sequence 4, Appli
10	146.6	12.5	1299	US-09-080-538-4	Sequence 4, Appli
11	146.6	12.5	1299	US-09-387-413-4	Sequence 4, Appli
12	146.6	12.5	1910	US-08-974-691-7	Sequence 7, Appli
13	146.6	12.5	2061	US-09-008-271A-16	Sequence 16, Appli
14	146.6	11.8	2038	US-08-631-097-7	Sequence 7, Appli
15	138.2	11.8	2038	US-08-810-712-11	Sequence 11, Appli
16	138.2	11.8	2465	PCT-US92-08080-1	Sequence 1, Appli
17	128.6	11.0	1329	US-09-705-448-4	Sequence 4, Appli
18	128.6	11.0	1329	US-09-705-448-4	Sequence 4, Appli
19	128.6	11.0	1353	US-08-974-691-1	Sequence 1, Appli
20	128.2	10.9	600	US-09-669-751-142	Sequence 142, App
21	110	9.4	1615	US-09-640-305-5	Sequence 5, Appli
22	110	9.4	1615	US-08-328-314-1	Sequence 1, Appli
23	109	9.3	2875	US-08-731-045-1	Sequence 1, Appli
24	109	9.3	2875	US-08-088-633-1	Sequence 1, Appli
25	97.8	8.3	2032	US-08-245-756-1	Sequence 1, Appli
26	97.8	8.3	2032	US-08-441-750-1	Sequence 1, Appli
27	97.8	8.3	2032	US-08-441-750-1	Sequence 1, Appli

28	97.8	8.3	2032	US-08-441-751-1	Sequence 1, Appli
29	97.8	8.3	2032	PCT-US92-02521-1	Sequence 1, Appli
30	95.8	8.2	2454	US-09-079-415-1	Sequence 1, Appli
31	74.8	6.4	392	US-09-702-705-1157	Sequence 1157, Ap
32	74.8	6.4	392	US-09-736-457-1157	Sequence 1157, Ap
33	67.6	5.8	663	US-09-280-116-265	Sequence 265, App
34	63.2	5.4	227	US-09-705-448-6	Sequence 6, Appli
35	57.6	4.9	264	US-09-016-434-292	Sequence 292, Appli
36	52.6	4.5	80	US-08-146-422-23	Sequence 23, Appli
37	52.6	4.5	80	US-08-146-422-24	Sequence 24, Appli
38	52.6	4.5	80	US-08-626-554-5	Sequence 5, Appli
39	52.6	4.5	80	US-08-626-554-29	Sequence 29, Appli
40	52.6	4.5	80	US-08-693-709-6	Sequence 6, Appli
41	51.6	4.4	80	US-09-230-590-4	Sequence 4, Appli
42	51.4	4.4	900	US-08-181-271A-4	Sequence 4, Appli
43	51.4	4.4	900	US-08-449-315-4	Sequence 4, Appli
44	51.4	4.4	900	US-08-444-803-4	Sequence 4, Appli
45	51.4	4.4	900	US-08-449-043-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-240-372-2
; Sequence 2, Application US/08240372
; Patent No. 5741665
; GENERAL INFORMATION:
; APPLICANT: KATO, ELIE K.
; APPLICANT: STUART, W. DORSEY
; TITLE OF INVENTION: LIGHT-REGULATED PROMOTERS FOR PRODUCTION
; TITLE OF INVENTION: OF HETEROLOGOUS PROTEINS IN FILAMENTOUS FUNGI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,372
; FILING DATE: 10-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 3918-0003.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-240-372-2

Query Match 76.5%; Score 897.2; DB 1; Length 1240;
Best Local Similarity 87.7%; Pred. No. 8.9e-310;
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY TCGTGTGCTGTACTACAGCTGTGTGATGATCCCGATTCCTCTGTACAAAGTAGTCTC 115
Db TCTTGCTCTCTCCAGGCGCTGTGTGATGATCCCGATTCCTGTACAAAGGAGTCTC 159
QY 116 TCCGTAAAGCGCTGAAGGAGATGAGACTTCTAGAAGACTTCTTGAGAAACAGTAGT 175
```

Db 160 TGAGGAGAGGCGCTGGAAGAGCATGGGCTTTTGAGAGACCTTCTGACAGAAACAGCATATG 219
 QY 176 GATCAGCAGCAAGTACTCCGGCTTGGTGAAGTTGCTAGCGTGCACCTTACCACTAC 235
 Db 220 GATCAGCAGCAAGTACTCCGGCTTGGGAGAGTGGCAGCGTGGCCCTGACCAACTAC 279
 QY 236 TTGATAGTCAATCTTTGGGAAGATCTACTCGGAAACCCCGCTCAAGATTCACCGTTC 295
 Db 280 TGGATAGTCAATCTTTGGGAAGATCTACTCGGAAACCCCGCTCAAGATTCACCGTTC 339
 QY 296 TCTTTGATAGTCAATCTTTGGGAGTTCCTGATCTGAGTTCCTCTATCTACTGCAAGAGATGCT 355
 Db 340 TGTTCACACTGGCTCTCTGACTTCTGGGTAACCTCTTATCTACTGCAAGAGATGCT 399
 QY 356 GCAAGAACCAACCAAGATTGATCCGAGAAAGTGTTCACCTTCCAGAACTTAGAGCAAC 415
 Db 400 GCAAAACCAACCAAGCTTGCACCCGAGAAAGTGTTCACCTTCCAGAACTTAGAGCAAC 459
 QY 416 CCTTGTCTATACCTACGTAAGGTAAGTGCATGCAAGAAATCTTAGAGCTTAGATCCGTCA 475
 Db 460 CCTTGTCTATACCTACGTAAGGTAAGTGCATGCAAGAAATCTTAGAGCTTAGATCCGTCA 519
 QY 476 CTGTCTCAACATTTGAGACATTCACAGACAGTAAAGACTTAGAACCCCAAGAGT 535
 Db 520 CTGTCTCAACATTTGAGACATTCACAGACAGTAAAGACTTAGAACCCCAAGAGT 579
 QY 536 ATGTCTTCAACCTTAGAGAAATTCGATGGAATCTTGTGTATGGAATACCATGCTGCGT 595
 Db 580 AGCTCTTCAACCTTAGAGAAATTCGAGGATCTCGGGAGATGGCTACCTCGCTGCGCT 639
 QY 596 CAGAGTACTGATACCTGTGTGTAACAACATGATGAACCGACACTAGTAGTCAAGACT 655
 Db 640 CAGAGTACTGATACCTGTGTGTAACAACATGATGAACCGACACTAGTAGTCAAGACT 699
 QY 656 TGTTCCTGAGTTTATGAGACAGGAATGGCCAGAGAGCATGCTGAGCTTGAAGTATTG 715
 Db 700 TGTTCCTGAGTTTATGAGACAGGAATGGCCAGAGAGCATGCTGAGCTTGAAGTATTG 759
 QY 716 ATCCATCTCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGACTAGTGC 775
 Db 760 ACCGCTCTACTACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACAGACTAGTGC 819
 QY 776 AATTCACTGTGACAGTGTGCATCAATGACCGGTGTGTTGATGCAATGGAAGTGTATGC 835
 Db 820 AGTTCACTGTGACAGTGTGCATCAATGACCGGTGTGTTGATGCAATGGAAGTGTATGC 879
 QY 836 AAGCTATCTTGAATACCGGTAAGTCCAAAGCTGTGCGGACCTAGCAGGAGATCTCAACA 895
 Db 880 AAGCTATCTTGAATACCGGTAAGTCCAAAGCTGTGCGGACCTAGCAGGAGATCTCAACA 939
 QY 896 TTTCAGCAAGCTATGAGACCAACAGAACAGTACGCTGATGTTGACATGATGTTGCA 955
 Db 940 TTCAGCAAGCTATGAGACCAACAGAACAGTACGCTGATGTTGACATGATGTTGCA 999
 QY 956 ACCTTGAAGTACAGCTGCTGTTTGAATCAACGCGCAAGTATCCACTGACCC 1015
 Db 1000 ACCTTGAAGTACAGCTGCTGTTTGAATCAATGCGCAAAATGTACCCACTGACCC 1059
 QY 1016 CCTCCGCTATACAGCAGGATCAAGGGTTCTGCACCAATGATTCCAAGTGAAGACC 1075
 Db 1060 CCTCCGCTATACAGCAGGATCAAGGGTTCTGTACCAATGCTTCAAGTGAAGATC 1119
 QY 1076 ATTCCAGAAATGATCTTTGGAGATGTGTTCAATCGTAGTACTACAGGCTCTTGA 1135
 Db 1120 ATTCCAGAAATGATCTTTGGAGATGTGTTCAATCGTAGTACTACAGGCTCTTGA 1179
 QY 1136 GGGCCAAACACTGTTGGGCTAGCTAAACCAATCTGA 1173
 Db 1180 GGGCCAAACACTGTTGGGCTAGCTAAACCAATCTGA 1217

RESULT 2

US-08-846-021A-6
 ; Sequence 6, Application US/08846021A
 ; Patent No. 5948682
 ; GENERAL INFORMATION:
 ; APPLICANT: Moloney, Maurice M.
 ; TITLE OF INVENTION: Preparation of Heterologous Proteins on
 ; TITLE OF INVENTION: Oil Bodies
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BERESKIN & PARR
 ; STREET: 40 King Street West
 ; City: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/846,021A
 ; FILING DATE: April 25, 1997
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gravelle, Micheline
 ; REGISTRATION NUMBER: 40,261
 ; REFERENCE/DOCKET NUMBER: 9369-039
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 364-7311
 ; TELEFAX: (416) 361-1398
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2733 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 850..1206
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1444..2729
 ; US-08-846-021A-6
 Query Match 76.3%; Score 894.8; DB 2; Length 2733;
 Best Local Similarity 88.4%; Pred. No. 1.2e-308;
 Matches 971; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
 QY 76 GCTGAGATCACCCGCACTTCTCTTACAAAGTAACTCTCCGTAAGGCGCTGAAGAA 135
 Db 1630 GCTGAGATCACCGAGATCTCTCTGTACAAAGGCAATCTCTGAGAAAGGCGTGAAGAG 1689
 QY 136 CATGAACTTCTAGAAAGCTTTCTTGCAAAACAGATATGGCATCGCAGCAAGTACTCC 195
 Db 1690 CATGAACTTCTAGAAAGCTTTCTTGCAAAACAGATATGGCATCGCAGCAAGTACTCC 1749
 QY 196 GCGTGGGTAAGTCTTACGCGGCACTTACCACTACTGATGTCAATCTTTGGG 255
 Db 1750 GCGTGGGTAAGTCTTACGCGGCACTTACCACTACTGATGTCAATCTTTGGG 1809
 QY 256 AAGATCTACCTCGGAACCCCGCTCAAGAGTTTCAACGTTCTCTTGTATCTGTTCCCT 315
 Db 1810 AAGATCTACCTCGGAACCCCGCTCAAGAGTTTCAACGTTCTCTTGTATCTGTTCCCT 1869
 QY 316 GACTTCTGGGTTCCCTCTATCTATCTGCAAGAGCAATGCTTCAAGAACCAAGATTC 375
 Db 1870 GACTTCTGGGTTCCCTCTATCTATCTGCAAGAGCAATGCTTCAAGAACCAAGATTC 1929
 QY 376 GATCCGAAAGAGTCTTCACTTCCAGAACTTAGGCAAAACCTTGTCTATACCTACGCT 435
 Db 1930 GATCCGAAAGAGTCTTCACTTCCAGAACTTAGGCAAAACCTTGTCTATACCTACGCT 1989

Db 1689 ATTCCAGAAATGATCTGGGGGATGTTTCATCCGAGATATTACAGCGTCTTGACA 1748

QY 1136 GGGCCAAACAACTCGTGGGTAGTAAAGCAATCTGA 1173

Db 1749 GGGCCAAACAACTCGTGGGTAGTAAAGCAATCTGA 1786

RESULT 4

US-08-318-193-15

; Sequence 15, Application US/08318193

; Patent No. 5641663

; GENERAL INFORMATION:

; APPLICANT: GARVIN, Robert T.

; APPLICANT: MALEK, Lawrence T.

; TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION

; TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY

; TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS

; NUMBER OF SEQUENCES: 91

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/318,193

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/935,314

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 18740/116 CACO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 832 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Other nucleic acid;

; DESCRIPTION: Synthetic DNA oligonucleotide

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 399..830

US-08-318-193-15

Query Match 17.4%; Score 203.6; DB 1; Length 832;

Best Local Similarity 75.1%; Pred. No. 4.5e-62;

Matches 254; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 57 CGTTGCTGTACTACAGCTGTGATGACCCGATTCCTCTACAAAGTAACTCT 116

Db 494 CGTTCCACCGCCGAAACGCTGCAAGATCACTAGATCCCATTTGACAGGGTAAGTCTT 553

QY 117 CGTAAAGCGCTGAAGGACATGACTTCTAGAAGACTTTGTCAGAAACAGATATGG 176

Db 554 GAGAAAGCGCTTGAAGAACACGCTTTGTTGAAGACTTTCTTGAAGAAAGCAACATACGG 613

QY 177 CATCAGCAGCAAGTACTCCGGCTTCGTTGTAAGTTGCTAGAGGTCCACTTACCACTACT 236

Db 614 TATCTCTCAAGTACTCTGGTTTGGTAAGTCGCTTCCTGTCATTTACCAACTACTT 673

QY 237 TGAATGATCAATCTTTGGAGATCTACCTCCGAACCCCGCTCAAGATTCACCGTCT 296

Db 674 GGACTCCCAATCTTCGTAAGATCTACTAGTACCCACCAAGAAATTCAGTCTCTT 733

QY 297 CTGTGATCTGTTCTCTGACTTCTGGGTTCCCTCTATCTACTGCAAGACATGCTTG 356

Db 734 GTTCGACACCGGTTCTTCTGACTTCGGGTCCATCGATTTACTGTAAGTCCACGCTTG 793

QY 357 CAAGAACCAACAAAGTTGATCCGAAAGTGTCCA 394

Db 794 TAAGAACCAACAAAGTTGATCCGAAAGTGTCTA 831

RESULT 5

5200327-6

; Patent No. 5200327

; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.

; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF

; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING

; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM

; STREPTOMYCES

; NUMBER OF SEQUENCES: 24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/224,568

; FILING DATE: 26-JUL-1988

; SEQ ID NO: 6:

; LENGTH: 828

5200327-6

Query Match 17.3%; Score 203; DB 6; Length 828;

Best Local Similarity 75.8%; Pred. No. 7.3e-62;

Matches 251; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 57 CGTTGCTGTACTACAGCTGTGATGACCCGATTCCTCTCTACAAAGTAACTCT 116

Db 494 CGTTCCACCGCCGAAACGCTGCAAGATCACTGAATCCCATTTGTAAGGGTAAGTCTT 553

QY 117 CGTAAAGCGCTGAAGGACATGAGACTTCTAGAAGACTTTGTCAGAAACAGATATGG 176

Db 554 GAGAAAGCGCTTGAAGAACACGCTTTGTTGAAGACTTTCTGCAAAAGCAACATACGG 613

QY 177 CATCAGCAGCAAGTACTCCGGCTTCGTTGTAAGTTGCTAGAGGTCCACTTACCACTACT 236

Db 614 TATCTCTCAAGTACTCTGGTTTGGTAAGTCGCTTCCTGTCATTTACCAACTACTT 673

QY 237 TGAATGATCAATCTTTGGAGATCTACCTCCGAACCCCGCTCAAGATTCACCGTCT 296

Db 674 GGACTCCCAATCTTCGTAAGATCTACTAGTACCCACCAAGAAATTCAGTCTCTT 733

QY 297 CTGTGATCTGTTCTCTGACTTCTGGGTTCCCTCTATCTACTGCAAGACATGCTTG 356

Db 734 GTTCGACACCGGTTCTTCTGACTTCGGGTCCATCGATTTACTGTAAGTCCACGCTTG 793

QY 357 CAAGAACCAACAAAGTTGATCCGAAAGTGTCCA 387

Db 794 TAAGAACCAACAAAGTTGATCCGAAAGTGTCCA 824

RESULT 6

5200327-5

; Patent No. 5200327

; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.

; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF

; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING

; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM

; STREPTOMYCES

; NUMBER OF SEQUENCES: 24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/224,568

; FILING DATE: 26-JUL-1988

SEQ ID NO:5:
LENGTH: 846
5200327-5

Query Match 17.0%; Score 199.8; DB 6; Length 846;
Best Local Similarity 76.2%; Pred. No. 1e-60;
Matches 246; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 74 CTGCTGAGATCAACCGCATTCCTCTCTACAAAGGTAAGCTCTCCGTAGAGGCGTGAAG 133
DB 517 CTGCAGAGATCACTGAATCCCATTTGACAGAGGTAAGCTTTGAGAAAGGCGCTTGAAG 576
QY 134 AACATGACCTTCTAGAGAGCTTCTTGACAGAAACAAGATGACATCAGACAGACTACT 193
DB 577 AACACGGTTTGTGTAAGAGCTTCTTGACAAAGCAAAATAGGATCTCTCCAGACTACT 636
QY 194 CCGGCTTGGTGAAGTTGCTAGCGTGCACCTTACCACTACTGATGATCAATCTTGG 253
DB 637 CTGGTTGGGTGAAGTGGCTGCTCCGTTCCATTCACCAACTTGAAGCTCCCAATCTTGG 696
QY 254 GGAAGATCTACCTGGAACCCGCGCTCAAGAGTTGACCGCTCTTTGATAGTGGTCTCT 313
DB 697 GTAAGATCTACTAGATCCCAACCAAGAAATTCACGTCTTGTTCGACACCGGTTCTT 756
QY 314 CTGACTTCTGGGTTCCCTCTATCTACTGCAAGAGCAATGCTGCAAGAACCAACCAAGAT 373
DB 757 CTGACTTCTGGGTTCCCATGATTTACTGTAAGTCCAGCGCTTGTAAAGAACCAACCAAGAT 816
QY 374 TCGATCCGAGAAAGTCTGACAC 396
DB 817 TCGACCAAGAAAGAGCTTACG 839

RESULT 7
US-08-318-193-13
Sequence 13, Application US/08318193
Patent No. 5641663

GENERAL INFORMATION:
APPLICANT: GARVIN, Robert T.
APPLICANT: MALEK, Lawrence T.
TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,193
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,314
FILING DATE:
APPLICATION NUMBER: US 07/224,568
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/116 CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:
LENGTH: 838 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid;
DESCRIPTION: Synthetic DNA oligonucleotide
FEATURE:
NAME/KEY: CDS
LOCATION: 399..836
US-08-318-193-13

Query Match 17.0%; Score 199.4; DB 1; Length 838;
Best Local Similarity 76.3%; Pred. No. 1.4e-60;
Matches 245; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 74 CTGCTGAGATCAACCGCATTCCTCTCTACAAAGGTAAGCTCTCCGTAGAGGCGTGAAG 133
DB 517 CTGCAGAGATCACTGAATCCCATTTGACAGAGGTAAGCTTTGAGAAAGGCGCTTGAAG 576
QY 134 AACATGACCTTCTAGAGAGCTTCTTGACAGAAACAAGATGACATCAGACAGACTACT 193
DB 577 AACACGGTTTGTGTAAGAGCTTCTTGACAAAGCAAAATAGGATCTCTCCAGACTACT 636
QY 194 CCGGCTTGGTGAAGTTGCTAGCGTGCACCTTACCACTACTGATGATCAATCTTGG 253
DB 637 CTGGTTGGGTGAAGTGGCTGCTCCGTTCCATTCACCAACTTGAAGCTCCCAATCTTGG 696
QY 254 GGAAGATCTACCTGGAACCCGCGCTCAAGAGTTGACCGCTCTTTGATAGTGGTCTCT 313
DB 697 GTAAGATCTACTAGATCCCAACCAAGAAATTCACGTCTTGTTCGACACCGGTTCTT 756
QY 314 CTGACTTCTGGGTTCCCTCTATCTACTGCAAGAGCAATGCTGCAAGAACCAACCAAGAT 373
DB 757 CTGACTTCTGGGTTCCCATGATTTACTGTAAGTCCAGCGCTTGTAAAGAACCAACCAAGAT 816
QY 374 TCGATCCGAGAAAGTCTGAC 394
DB 817 TCGACCAAGAAAGAGCTTCTA 837

RESULT 8
US-09-032-523-6
Sequence 6, Application US/09032523
Patent No. 6232454

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guejler, Karl
APPLICANT: Baugh, Mariah
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANTU01
CLONE: 1515165
US-09-032-523-6

Query Match 16.4%; Score 192.2; DB 3; Length 2073;
Best Local Similarity 56.5%; Pred. No. 1e-57;
Matches 406; Conservative 0; Mismatches 298; Indels 15; Gaps 2;

```
QY 87 CCGCATTCCTCTCTACAAAGTAAGTCTCCGTAAGGCGCTGAAGAGATGGAATCTT 146
DB 104 CAGGGTCCCTCAGAGGATCCGTCTTCAAGAGAGCTGGGACAGGACGACT 163
QY 147 AGAAGCTTCTTGCAGAAACAGATGAGATCA-----GCAGCAAGTACTCCGCTT 200
DB 164 CTCGAGTTCTGGAATCCATATATTGACATGATCAGTTACCGAGTCTGCTCAAT 223
QY 201 CGGTGAAGTTCTAGAGCTGCACTTACCACTACCTTGATAGTCAATGCTTTGGAAAT 260
DB 224 GGACCGAAGGCCAAGAACCCCTCATCAACTTGTGATATGGAATATCTTCGCACTAT 283
QY 261 CTACCTCGAAGCCCGCTCAAGAGTACCGTTCTTTGATATGAGTTCTCTGACTT 320
DB 284 CTCATTGCTCCCTCCACACAGACTTCTGATCTTCTGACACTGGCTCTTCAACT 343
QY 321 CTGGGTTCCCTCTATCTACTGCAAGAGCAATGCTGCAAGAACCAAGATTCATCC 380
DB 344 CTGGGTCCTCTGTGTATGCACTAGCCAGCTGCAAGACGACAGGTTCCAGCC 403
QY 381 GAGAAAGTCTCCACTCTTCCAGAACTTAGCGAAACCTTGTCTATATACATACGATACG 440
DB 404 TTCCAGTCCAGACATACAGCCAGCCAGGTCAATCTTCTCCATTCAGATGGAACCG 463
QY 441 TAGCATGCAAGAACTTAGGCTATGATACCGTACGTCCTCCAACTTGTGACATTTCA 500
DB 464 GAGCTTGTCCGGATCATGTGAGCCGACCAAGTCTGTGGAAGAACTAAACGCTTGG 523
QY 501 ACAGACGATGAGCTTAGCAACCAAGAGATGATGTTCTTCACTATGCAAGATTTCA 560
DB 524 CCAGCAAGTTTGGAGAAAGTGTCAAGAGCCAGGCAACCTTTGTGATGCAAGTTTGA 583
QY 561 TGGCATCTTGTGATGGAATACCATGCTGCGGTCAAGATCTCGATACCTGTGTTGA 620
DB 584 TGAATTTCTGGGCTTGGAATACCTCTCTTGTGTGGAGAGATGACTCCAGATTTGA 643
QY 621 CAACATGATGAACGACACCTAGTAGTCAAGACTTGTCTCGGTTTAACTGACAGAA 680
DB 644 CAACATGATGCTAGAACCTGTGGAGCTTGGCCGATTTTGTCTAAGAGCACTAA 703
QY 681 -----TGGCAGAGAGGATGCTCAGCTTGTGAGCTATTGATCATCTTACTAC 731
DB 704 CCAGAAAGGTGTGCGGAGGAGGAGCTGATTTTGGAGGCTACAGCACTCCATTTCTC 763
QY 732 AGGATCTTCTACTGGGTTCAGTCACTGTGACAGCACTACTGCAATTCAGTGTGACA 790
DB 764 TGGAGGCTGAATTTGGGTCCAGTCAACAGCAAGCTTACTGGGAGATGCACTGTGATA 822
```

RESULT 9

US-08-723-938-4
Sequence 4, Application US/08723938
Patent No. 5776759
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,938
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LINCNOT02
CLONE: 312099
US-08-723-938-4

Query Match 12.5%; Score 146.6; DB 1; Length 1299;
Best Local Similarity 50.4%; Pred. No. 1.4e-41;
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;

```
QY 216 CGTGCACTTACCACTACTGTGATGCAATATCTTTGGAGAGATCTACTCGGAACCC 275
DB 217 CGTACCTCTCTGAACTTACAGGAGATGTGCAATTTTGGGAAATTTGGGCTGGGAACGC 276
QY 276 GCTTCAAGATTCACCGTTCTCTTATATATGTTTCTGTACTTGTGGTTCCCTCTA- 334
DB 277 TCCACAAACTTCACTGTGCTTTGACACTGCTCTCCAACTCTGTGGTCCCGTCCAG 336
QY 335 -----TCTACTGCAAGCAATGCTGCAAGAACCAACCAAGATTCGATCCGAAAGTC 389
DB 337 GAGATGCCACTTTTCAAGTGTGCTGCTGCTGTTACACCAAGTTTGAATCCCAAAGCTTC 396
QY 390 GTCCACTTTCAGAACTTAGCAAAACCTTGTCTATATACATAGGTACAGTACATGCA 449
DB 397 TACTCTCTTCAGGCGCAATGGGAGCAAGTTTGTGCATTCAATATGAACTGGGCGGTAGA 456
QY 450 AGGAATCTTAGGCTATGATACCGTCACTGTCTTCCAACTTTGGGACATTTCAAGACAGT 509
```

```

Db      457 TGGAAATCTGAGGAGGAGCAAGCTGATCTATTGGTGGATCAAGGCTGATCATGATTTT 516
Qy      510 AGGACTTAGACCCCAAGAACAGGTGATGTCTTCACTTATGAGATTTGATGATCTCT 569
Db      517 CGGGGAGGCTCTCTGGAGGCCAGCTGTCTTCCGCTTTTGGCCATTTTGAAGGATATT 576
Qy      570 TGGTATGGCATACCCATCGCTCGCTCAGAGTACTGATACCTGTGTGTTGACAAATGAT 629
Db      577 GGGCCTCGGTTTTTCCCATTTCTGTGTGGAAGAGTGGGCCCCGATGATGATCTGT 636
Qy      630 GAACCGACACTGATGACTCAAGACTTGTCTCGGTTTACATGACAGGAATGGCCAGGA 689
Db      637 GGAGCAGGGGCTATTGATTAAGCTGTCTCTCTTTTACCTCAACAGGAGCCCTGAAGA 696
Qy      690 G-----AGCATGCTCAGGCTTGAGGCTATTGATTCATCTCTACAGAGATCTCT 740
Db      697 GCTTATGAGAGAGAGCTGTGTCTTGGGGGCTCGGACCCGGACACTATATCCACCCCT 756
Qy      741 TCACCTGGGTTCCAGTCACTGTGTCAGCAGTACTGGAATTCAGTGTGACAGTGCACCAT 800
Db      757 CACCTTGGTCCAGTCAAGGTCCTCCGCTTACCTGAGAGATCCATGAGAGCTGTGAAGT 816
Qy      801 CAGCGGTGTGTTGTGATGTAAGGTGATGTCAGTACTATCTTGATATCCGTTACGTC 860
Db      817 GGGCCCAAGGCTGACTCTCTGTGCAAGGGCTGTGCTGCATCTGGAATACGGGACGTC 876
Qy      861 CAAGCTGTGCGGACCTAGCAGGACATTTCTCAACATGCAAGATATTGAGCCACACA 920
Db      877 CTTCAATCAGAGACCCACTGAGAGATCCGGGCCCTGACATGACGATTTGGGGAAATCC 936
Qy      921 GAACCACTACGCTGATTTGATAGATGCGACAACCTTACATGATGCTTACAGTTGT 980
Db      937 CTGTGCTGGTGGAGTATCATCTGTGCTGGAATTCGAAAGCTCCCGCAGTCTC 996
Qy      991 CTTTGAATCAACGGCAAGATGTACCACTGACCCCTTCGCGCTATACCAAGCCAG 1035
Db      997 CTTCCTTCTGGGGGGTCTGTTTAACCTCAAGGCCCATGATTAAGTATCAG 1051

```

RESULT 10

```

; Sequence 4, Application US/09080538
; Patent No. 5965129
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Coleman, Roger
; TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,538
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/723,938
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0125 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

```

```

; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; IMMEDIATE SOURCE:
; LIBRARY: LUNGMOT02
; CLONE: 312099
; US-09-080-538-4

Query Match      12.5%; Score 146.6; DB 2; Length 1299;
Best Local Similarity 50.4%; Pred. No. 1.4e-41;
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;

Qy      216 CGTGCACCTTACCAACTATCTGATGCAATTACTTTGGAAAGATCTATCGGAACCC 275
Db      217 CGTACTCTCTGAACTACAGGATGTGCAATTTTGGGAAATTTGGGCTGGGAACGCC 276
Qy      276 GCTTCAAGATTCACCGTCTCTTTGATATGTTGTTCTCTGACTTCTGGGTTCCCTTA- 334
Db      277 TCCACAAATCTCATGTTGCTTTGACATGGCTCTTCAATCTTGGGTCCTGCTCAG 336
Qy      335 -----TCTACTGCAAGAGCAATGCTGCAAGAACCAACCAAGATTTCATCCGAAAGTC 389
Db      337 GAGATGCCACTTCTTCACTGATGCTGCTGCTGTATACACACGATTTTATCCAAAGCCTC 396
Qy      390 GTCCACCTTCAGAACTTAGGCAAACTTGTCTATATACATACTAGGTACAGTATGATGCA 449
Db      397 TAGTCTCTTCAGGCAATGAGACCAAGTTTGCCATTAATATGGAATCGGGCGGATGA 456
Qy      450 AGGATCTTAGGCTATGATACCGTCACTGTCTCAACATTTGTGACATTTCAAGACAGT 509
Db      457 TGGATCTTGAAGCGAGGCAAGCTGACTATGTGGAATTCAGAGGTGATCATGATATT 516
Qy      510 AGGACTTAGACCCCAAGAACAGGTGATGTCTTCACTATGAGAAATTCATGATGATCTCT 569
Db      517 CGGGAGGCTCTCTGGAGGCCAGCTGTGCTTCTTTCCTTTCCTTATTTGAGGATATT 576
Qy      570 TGGTATGCAATACCCATGCTCGGTCAAGTACTGATACCTGTGTTGACAAATGAT 629
Db      577 GGGCCTCGGTTTTTCCCATTTCTGTGTGGAAGAGTTGGGCCCCGATGATGATCTGT 636
Qy      630 GAACCGACACTGATGCTCAAGACTTGTCTCGGTTTACATGACAGGAATGGCCAGGA 689
Db      637 GGAGCAGGGGCTATTGATTAAGCTGTCTCTCTTTTACCTCAACAGGAGCCCTGAAGA 696
Qy      690 G-----AGCATGCTCAGGCTTGAGGCTATTGATTCATCTCTACAGAGATCTCT 740
Db      697 GCTGATGAGAGAGAGCTGTGTCTTGGGGGCTCGGACCCGGACACTATATCCACCCCT 756
Qy      741 TCACCTGGGTTCCAGTCACTGTGTCAGCAGTACTGGAATTCAGTGTGACAGTGCACCAT 800
Db      757 CACCTTGGTCCAGTCAAGGTCCTCCGCTTACCTGAGAGATCCATGAGAGCTGTGAAGT 816
Qy      801 CAGCGGTGTGTTGTGATGTAAGGTGATGTCAGTACTATCTTGTGATACCGTTACGTC 860
Db      817 GGGCCCAAGGCTGACTCTCTGTGCAAGGGCTGTGCTGCATCTCGATACGGGACGTC 876
Qy      861 CAAGCTGTGCGGACCTAGCAGGACATTTCTCAACATTCAGAGATATTGAGCCACACA 920
Db      877 CTTATACAGAGACCCACTGAGAGATCCGGGCCCTGATGACAGCATTTGGGGAAATCC 936
Qy      921 GAACCACTACGCTGATTTGATAGATGCGACAACCTTACATGATGCTTACAGTTGT 980
Db      937 CTGTGCTGGGAGTATCATCTGTGTCTGGAATTCGAAAGCTCCCGCAGTCTC 996

```

QY 981 CTTTGATGATCAAGGCAAGTATGATCCAGCTGACCCCTCCGCTATACGACG 1035
DB 997 CTTCTCTTGGGGGGGTCTGTGTTTAACTCAAGGCGCCAGATGATGATCAACG 1051

RESULT 11

US-09-387-413-4
; Sequence 4, Application US/09387413
; Patent No. 6475485

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
Coleman, Roger

TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/387,413
FILING DATE: 31-Aug-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/080,538
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0125 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: <Unknown>
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGNOT02
CLONE: 312099

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-387-413-4

Query Match 12.5%; Score 146.6; DB 4; Length 1299;
Best Local Similarity 50.4%; Pred. No. 1.4e-41;
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;

QY 216 CGTGCCACTTACCACTACCTGATAGTCAATCTTGGGAAGATCTACCTGGAAACCC 275
DB 217 CTTACCTCTCTCGAATCAAGGATGCACTATTTTGGGAAATTGGGCTGGGAAGCC 276
QY 276 GCGTCAAGAGTCAACCGTCTCTTTGATGATGTTCTCTGACTTCTGGGTTCCCTTA- 334
DB 277 TCACAAAACCTCACTGTTCCTTGAACACTGGCTCTCCAAATCTCGGGTCCGTCAG 336
QY 335 -----TCTATTGCAAGGCAATGCTTGCAAGAACCAAGATTCATCCGAAAGTC 389

DB 337 GAGATGCCACTTCTCAGTGTGCGCTGCTGTATACACCAAGATTTGATCCAAAGCTC 396
QY 390 GTCCACCTTTCAGAACTTAGGCAAACTTGTCTATPACCTACGTTACAGTACATGCA 449
DB 397 TAGCTCTCTTCCAGGCAATGAGGACCAAGTTTCCATTCATPAGAACTGGGCGGTAGA 456
QY 450 AGGAATCTTAGGCTATGATACCGTCACTGTCCAAACATGTGACATTCAACAGCACT 509
DB 457 TGGATCTTGAGCGAGAGACAGCTGATTTGTGGAATCAAGGTGATCACTAGTATTTT 516
QY 510 AGGACTTACACCCCAAGCAAGAGTGTCTTTCACCTTACAGTATGCAATTCATGCTCT 569
DB 517 CGGGAGAGCTCTCTGGAGACCCAGCTGTCTTCTGCTTTGGCCATTTTGGATGATAT 576
QY 570 TGGTATGCAATCAATCCATCGCTCGCTCAGAGTACTGATACCTGTGTGTAACAATGAT 629
DB 577 GGGCTCTGGTTTCCCATTTCTGTGTGGAAGAGTTCGGCCCGATGATGATCTGTGT 636
QY 630 GAACGACACCTAGTATGCTCAAGACTTGTCTCGGTTTACATGACAGGAATGGCCAGGA 689
DB 637 GGAGCAGGGGCTATTGATTAAGCTGTCTTCTCTTTTACCTCAACAGGACCCGTAAGA 696
QY 690 G-----AGCATGCTCAGCGCTTGGAGCTATTGATCCATCTACTACAGATCTCT 740
DB 697 GCGTATGAGAGAGAGCTGTCTGTGGGGGCTCGACCCGGACACTATCCACCCCT 756
QY 741 TCATCGGTTTCAAGTCACTGTGCAAGATGTCGGAATTCAGTGTGACAGTGCAT 800
DB 757 CACCTTCGTCAGATCAAGCTGCCGCTACATGCGAGATCCACATGACGTTGTAAGT 816
QY 801 CAGCGTGTGTGTGTGATGATGGAAGTGAATGTCAAGTATCTTGGAATACCGGTACGT 860
DB 817 GGGCCCAAGGCTGACTCTGTGTGCAAGGGCTGTGTGATCTCGAATACGGGACGTC 876
QY 861 CAAGCTGTGCGACCTAGCAGGACATCTCAACATTCAGCAAGTATGAGCCACAGA 920
DB 877 CCGATCAAGAGACCCATGAGAGATCCGGGCCCTCGATGACGATTTGGGGAATCCC 936
QY 921 GAACGATACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 980
DB 937 CTTCTCTTGGGGGAGTATCATCTCTGTGTGCAAGATCCCAAGCTCCCGCAAGTCTC 996

RESULT 12

US-08-974-691-7
; Sequence 7, Application US/08974691
; Patent No. 6225103

GENERAL INFORMATION:

APPLICANT: Keoloch, Gerald
APPLICANT: Lin, Xini

TITLE OF INVENTION: Cloning and Characterization of Napsin

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 W. Peachtree
STREET: St.
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,691

FILING DATE: 20-NOV-1997


```
QY 276 GCCTCAAGATTGACCGTCTCTTTGATCTGCTTCTGACTTCTGGGTCCCTCTA- 334
| | | | |
DB 971 TCCACAAAATTCTCACTGTTCCTTTGACACTGGCTCTCCAACTCTCGGGTCCGTCAG 1030
| | | | |
QY 335 -----TCTACTGCAAGAGCAATGCTGCAAGAACCAACAAAGATTGCAAGAAAGTC 389
| | | | |
DB 1031 GAGATGCACTTCTTCAAGTGTGCTGCTGCTGCTTACACACCGATTGATCCCAAGCTTC 1090
| | | | |
QY 390 GTCCACCTTCGCAAACTTAGGCAAAACCTTGTCTATACACTACGCTACAGTACATGCA 449
| | | | |
DB 1091 TAGCTCTCTCCAGGCAATGGAGACCAAGTTTGCATTCAATATGAACTGAGCGGGTAGA 1150
| | | | |
QY 450 AGGAATCTTAGGCTATATACCGTCACTGTCTCCAAATTGTGACATTCAACAGACAT 509
| | | | |
DB 1151 TGAATCTTAGGAGAGCAAGCTGATATGTGTGAAATCAAGGGTGCATCAGTAAATTT 1210
| | | | |
QY 510 AGGACTTAGACCCCAAGAACCAAGTGTATGTTTCACTATGCAATTTGATGGACATCT 569
| | | | |
DB 1211 CGGGAGAGCTCTCTGGAGAGCCCAAGCTGTGTCTTGCTTTGCCATTTTGAATGGATATT 1270
| | | | |
QY 570 TGGTATGCAATACCATCGCTCGGTGCAAGTACTGATACCTGTGTTTGAACAATGAT 629
| | | | |
DB 1271 GGGCTCTGCTTCCCACTCTGTCTGTGAAGAGTTCGGCCCCGATGATGATGCT 1330
| | | | |
QY 630 GAACCGACACTTACTGATCAAGACTTGTCTCGGTTTACATGAGACAGAAAGCCAGGA 689
| | | | |
DB 1331 GAGACAGGGGCTATGATTAAGCTGTCTTCTCTTTTACCTCAACAGGAAACCTGAAAG 1390
| | | | |
QY 690 G-----AGCATGCTCAAGCTTGAGCTATGATTCATTCCTACTACAGAGATCTCT 740
| | | | |
DB 1391 GCCTGATGAGAGAGAGCTGTGCTCTGGGGGGCTGGACCCGGACACTACATCCACCCCT 1450
| | | | |
QY 741 TCATGCGGTTCCAGTCACTGTGACAGATACCTGGCAATTCACTGTGACAGTGTCACT 800
| | | | |
DB 1451 CACTTCTGTGCAAGTCAAGGCTCTGCTGCTTACTGGCAATCAATGAGAGTGTGAAGT 1510
| | | | |
QY 801 CAGCGGTGTGTTTGTGATGTGAAGGTGATGTCAAGCTATCTTGAATACCGTATGCT 860
| | | | |
DB 1511 GGGCCCAAGGGCTGACTCTGTGTGCAAGGGGTGTGTCATCTGTGATACGGGCACTGTC 1570
| | | | |
QY 861 CAAGCTGTGAGACTGACGACGACGACATTTCAATTCAGACATTCATGAGCTATGAGCCACA 920
| | | | |
DB 1571 CTTGATACAGAGACCCACTGAGAGATCCGGGCTCTGATGACACCAATGGGGAAATCCC 1630
| | | | |
QY 921 GAAACGATACGATGATTTGACATAGATTGACACCACTTACCTTACATGCTTACATGTTGT 980
| | | | |
DB 1631 CTTGCTGTGGGAGATGATCATCTGTGCTGGAATCCCAAGCTCCCCGCAATCTC 1690
| | | | |
QY 981 CTTTGAATCAACGCAAGATGATCCCACTGACCCCTCGGCTTATACCAAGCCAG 1035
| | | | |
DB 1691 CTTCTTCTTGGGGGGGTCTGTGTTTAACTTCAAGGCCCATGATTAATGCTATCCAG 1745
| | | | |
RESULT 14
US-09-705-448-2
; Sequence 2, Application US/09705448
; Patent No. 6432690
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong
; APPLICANT: Bruno, Sandra A.
; APPLICANT: Elsenboss, Laura A.
; APPLICANT: Fogliano, Michael
; APPLICANT: Cohen, Victoria L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: HUMAN ASPARTIC PROTEASES
; FILE REFERENCE: PR-0458-1 CIP
; CURRENT APPLICATION NUMBER: US/09/705,448
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/116,641
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/008,271
; PRIOR FILING DATE: 1998-01-16
```

```
NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
; FEATURE:
; OTHER INFORMATION: 877617, LUNGAST01
US-09-705-448-2
```

```
Query Match 12.5%; Score 146.6; DB 4; Length 2061;
Best Local Similarity 50.4%; Pred. No. 2e-41;
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;
```

```
QY 216 CGTGCACTTACCAATCACTGATGATGATCAATCTTGGAAAGATCTACCTGGAACCC 275
| | | | |
DB 911 CGTACCTCTCTGAACTTACAGGATGTGACGATATTTTGGGAAATTTGGGCTGGGAACGCC 970
| | | | |
QY 276 GCCTCAAGAGTTCAACCGTCTCTTTGATATGTTTCTCTGACTTCTGGGTTCCCTCTA- 334
| | | | |
DB 971 TCCACAAAATTCTCACTGTTCCTTTGACACTGGCTCTCCAACTCTCGGGTCCGTCAG 1030
| | | | |
QY 335 -----TCTACTGCAAGAGCAATGCTGCAAGAACCAACAAAGATTGCAAGAAAGTC 389
| | | | |
DB 1031 GAGATGCACTTCTTCAAGTGTGCTGCTGCTGCTTACACACCGATTGATCCCAAGCTTC 1090
| | | | |
QY 390 GTCCACCTTCGCAAACTTAGGCAAAACCTTGTCTATACCTACGCTACAGTACATGCA 449
| | | | |
DB 1091 TAGCTCTCTCCAGGCAATGGAGACCAAGTTTGCATTCAATATGAACTGAGCGGGTAGA 1150
| | | | |
QY 450 AGGAATCTTAGGCTATATACCGTCACTGTCTCCAAATTGTGACATTCACACAGACAT 509
| | | | |
DB 1151 TGAATCTTAGGAGAGACAGCTGACTATGTGTGAATCAAGGTGCATCAGTAAATTT 1210
| | | | |
QY 510 AGGACTTAGACCCCAAGAACCAAGTGTATGTTTCACTTACGAGAAATTCGATGGACATCT 569
| | | | |
DB 1211 CGGGAGAGCTCTCTGGAGAGCCCAAGCTGTGTCTTGCTTTGCCATTTTGAATGGATATT 1270
| | | | |
QY 570 TGGTATGCAATACCATCGCTCGGTGCAAGTACTGATACCTGTGTTTGAACAATGAT 629
| | | | |
DB 1271 GGGCTCTGCTTCCCACTCTGTCTGTGAAGAGTTGCGCCCCGATGATGATGCT 1330
| | | | |
QY 630 GAACCGACACTTACTGATCAAGACTTGTCTCGGTTTACATGAGACAGAAAGCCAGGA 689
| | | | |
DB 1331 GAGACAGGGGCTATGATTAAGCTGTCTTCTCTTTTACCTCAACAGGAAACCTGAAAG 1390
| | | | |
QY 690 G-----AGCATGCTCAAGCTTGAGCTATGATTCATTCCTACTACAGAGATCTCT 740
| | | | |
DB 1391 GCCTGATGAGAGAGAGCTGTGCTCTGGGGGGCTGGACCCGGACACTACATCCACCCCT 1450
| | | | |
QY 741 TCATGCGGTTCCAGTCACTGTGACAGATACCTGGCAATTCACTGTGACAGTGTCACT 800
| | | | |
DB 1451 CACTTCTGTGCAAGTCAAGGCTCTGCTGCTTACTGGCAATCAATGAGAGTGTGAAGT 1510
| | | | |
QY 801 CAGCGGTGTGTTTGTGATGTGAAGGTGATGTCAAGCTATCTTGAATACCGTATGCT 860
| | | | |
DB 1511 GGGCCCAAGGGCTGACTCTGTGTGCAAGGGGTGTGTCATCTGTGATACGGGCACTGTC 1570
| | | | |
QY 861 CAAGCTGTGAGACTGACGACGACGACATTTCAATTCAGACATTCATGAGCTATGAGCCACA 920
| | | | |
DB 1571 CTTGATACAGAGACCCACTGAGAGATCCGGGCTCTGATGACACCAATGGGGAAATCCC 1630
| | | | |
QY 921 GAAACGATACGATGATTTGACATAGATTGACACCACTTACCTTACATGCTTACATGTTGT 980
| | | | |
DB 1631 CTTGCTGTGGGAGATGATCATCTGTGCTGGAATCCCAAGCTCCCCGCAATCTC 1690
| | | | |
QY 981 CTTTGAATCAACGCAAGATGATCCCACTGACCCCTCGGCTTATACCAAGCCAG 1035
| | | | |
DB 1691 CTTCTTCTTGGGGGGGTCTGTGTTTAACTTCAAGGCCCATGATTAATGCTATCCAG 1745
| | | | |
```

```
RESULT 15
US-08-631-097-7
```

```

Sequence 7, Application US/08631097
Patent No. 5968816

GENERAL INFORMATION:
APPLICANT: Kimchi, Adi
TITLE OF INVENTION: Tumor Suppressor Genes,
TITLE OF INVENTION: Protein Encoded Thereby, and Use of Said Genes and Protein
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wigman, Cohen, Leitner, & Myers, P.C.
STREET: 900 17th Street, N.W., Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/631.097
FILING DATE: 12-Apr-96
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/11598
FILING DATE: 12-Oct-94
ATTORNEY/AGENT INFORMATION:
NAME: Cohen, Herbert
REGISTRATION NUMBER: 25,109
REFERENCE/DOCKET NUMBER: 0744.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)463-7700
TELEFAX: (202)473-6915
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: No
FRAGMENT TYPE: No. 5968816 applicable
ORIGINAL SOURCE:
ORGANISM: homo sapiens
STRAIN: not applicable
INDIVIDUAL ISOLATE: not applicable
DEVELOPMENTAL STAGE: not applicable
HAPOTYPE: not applicable
TISSUE TYPE: blood
CELL TYPE: Leucocyte
CELL LINE: HeLa
ORGANELLE: not applicable
IMMEDIATE SOURCE:
LIBRARY: not applicable
CLONE: not applicable
POSITION IN GENOME:
CHROMOSOME/SEGMENT: not applicable
MAP POSITION: not applicable
UNITS: not applicable

FEATURE:
NAME/KEY: This is the DNA sequence
NAME/KEY: claimed in 15(vi) as the Cathepsin gene in Fig. 15.
LOCATION: not available
IDENTIFICATION METHOD: experiment-
IDENTIFICATION METHOD: in specification
OTHER INFORMATION: prevention of IFN-2
OTHER INFORMATION: promoted cell death
PUBLICATION INFORMATION: not available
US-08-631-097-7

```

	Matches	443;	Conservative	0;	Mismatches	378;	Indels	48;	Gaps	3
Qy	216	CGTGCACCTTAACCACTACCTTGAATGATCACTACTTTGGGAAAGATCTACCTCGGAACCCC	275							
Db	255	CGAGTGCTCAAGAACTACATATGACGCCCACTACTACGGGGAGATTGGCATCGGACGCC	314							
Qy	276	GCCCAAGAGTGCACCGTCTCTTGTGADACTGGTGTCTCTGACTTCTGGGTTCCTCTAT	335							
Db	315	CCCCAAGTCTTACACAGTCGTCTTTCGACAGGGCTCTTCCAACTGTGAGGTCCCTTCAT	374							
Qy	336	CTACTGCAA-----GAGCAATGCGCTGCAGAAACACCAAAAGATTGCATCCGAAAGTTC	389							
Db	375	CCAATTGCAACCTGTGGACATCGTGTGTGATGCCAACAAAGTAAACAACAGACAAACATC	434							
Qy	390	GTCACCTTCCAGAACTTAGGCAAAACCTTGTCTATACACTACGGTACAGGTAGATGCA	449							
Db	435	CAGCACTTACGTGAAGATGGTACCTCGTTTGACATCCACTATATGCTTCGGGACAGCTTTC	494							
Qy	450	AGGAATCTTAGGCTATATATACCGTCACTGTCTCAAC-----	486							
Db	495	CGGGTACCTTAGCCAGACACTGTGTGGTGCCCTGCAAGTCAGCGTCGACGCTCTGC	554							
Qy	487	-----ATTGTGACATTTCAACAGACAGTATGSACTTAGCACCCAAAGAACAGGTGA	536							
Db	555	CTGTGGGGGTGTCAAATGATGAGAGGACAGTCTTTGGGGAGGCCACCAAGACAGCCAGGAT	614							
Qy	537	TGTTCTTCACTTANGCAGAAATTCAGTGGCATCTTGTGATGGCATACCCATCGCTCGCTC	596							
Db	615	CACCTTATCGACGCCAAGTTGATGTGCATCTTGGGCATGGCTTACCCCGCATCTCCGT	674							
Qy	597	AGATTAATCCATACCTGTGTGTCACATGATATGAACGACACTAGTAGTCTAAAGATT	656							
Db	675	CAACAAAGGTGTGCCGCTTTCGACCAACCTGATACAGAAAGCTGTGTGACCAAGAAAT	734							
Qy	657	GTTCCTCGATTTAACATGACAGGAA-----TGGCCAGAGAGACATGTCTCAGCTTGG	707							
Db	735	CTTCTCTCTTCTACTGTGACAGGAGACCCAGATGCGGACCTGGGGGTAGCTGATGTCTGG	794							
Qy	708	AGCTATTGATCCATCTACTACACAGGATCTTTCACTGGGTTCAGTCACTGTGCAGCA	767							
Db	795	TGGACACAGACTCCAAATATTACAGAGGGTCTGTGCTTACTGATATGCCGCCGACAGC	854							
Qy	768	GTAATGGCAATTAAGTGTGACAGTGTACCATTAAGCGGTGTGGTGTGTGATGTGAAG	827							
Db	855	CTACTGGCAGGTTCACCTGACCCAGGTGAGGTGACGACGCGGGCTGACCCTGTGTAAAGA	914							
Qy	828	TGATATGCAAGCTATCTTGGATACCGGTACGTCMAAGTGTGCGAACCCTAGCAGGACAT	887							
Db	915	GGGCTGTGAGGCCATTTGTGACACAGGACCTTCTCTATGTGTGGCCCTGGTGTGATGAGT	974							
Qy	888	TCTCAACATTCAGCAAGCTATTGGAGCCACACAGAACCAAGTACGATGTTTGAATGAA	947							
Db	975	GCGGAGCTGTGAAGAGGCATCGGGGCGGTGCGCTGATTAAGGGCAGATACATGATCCC	1033							
Qy	948	TTGCGACAACCTTAGCTACATGCTCTACAGTTGTCTTTGAGATCAACGCGACAGATGATCCC	1007							
Db	1035	CTGTGAAGAAGGTGTCCACCCTGCGCGGATCACACTGAAGCTGGGAGGCAAGGCTACAA	1093							
Qy	1008	ACTGACCCCTCCGCGCTATACCAACCCAGG	1036							
Db	1095	GCTGTCTCCCAAGAGACTACAGCTCAAGG	1123							

Search completed: February 5, 2004, 05:20:15
Job time : 102 secs

Query Match	11.8%;	Score 138.2;	DB 2;	Length 2038;
Best Local Similarity	51.0%;	Pred. No. 2e-38;		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 03:08:18 ; Search time 2406 Seconds
(without alignments)
1795.886 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173
Sequence: 1 atgactctcctaagctctt.....ggctagctaagcaatctga 1173

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894.8	76.3	2733	11	US-09-893-525-6
2	360.8	30.6	1136	12	US-09-971-392-41
3	359.2	30.6	1131	12	US-10-074-978A-7
4	359.2	30.6	1688	12	US-10-074-978A-3
5	357.6	30.5	1131	12	US-10-074-978A-5
6	357.6	30.5	1131	12	US-10-074-978A-11
7	356	30.3	1131	12	US-10-074-978A-9
8	253	21.6	2201	13	US-10-295-027-178
9	248.4	21.2	1366	10	US-09-969-708-441
10	248.4	21.2	1366	10	US-10-101-510-457
11	248.4	21.2	1366	13	US-09-873-319-641
12	248.4	21.2	1366	13	US-09-960-706-984
13	230.6	19.7	1270	12	US-10-051-874-53
14	226.8	19.3	1173	12	US-10-274-639-30

16	150.6	12.8	1341	13	US-10-220-083-1	Sequence 1, Appli
17	149.2	12.7	1530	10	US-09-887-576-795	Sequence 795, App
18	146.6	12.5	1299	14	US-10-094-080-4	Sequence 4, Appli
19	146.6	12.5	1365	15	US-10-288-222A-29	Sequence 29, Appli
20	146.6	12.5	1425	10	US-09-964-899-10	Sequence 10, Appli
21	146.6	12.5	2061	13	US-10-180-719-16	Sequence 16, Appli
22	144.8	12.3	1521	10	US-09-938-842A-2245	Sequence 2245, Ap
23	144.8	12.3	1521	12	US-09-938-842A-2245	Sequence 13, Appli
24	141.6	12.1	1360	9	US-09-789-919-43	Sequence 43, Appli
25	141.6	12.1	1360	9	US-10-168-425-27	Sequence 27, Appli
26	138.8	11.8	1234	15	US-10-168-425-27	Sequence 27, Appli
27	138.2	11.8	2038	13	US-10-101-510-25	Sequence 25, Appli
28	138.2	11.8	2076	15	US-10-084-817-172	Sequence 172, App
29	138	11.8	1828	13	US-10-339-351-5	Sequence 5, Appli
30	136.6	11.6	1988	13	US-10-007-926A-128	Sequence 128, App
31	129.4	11.0	1194	15	US-10-168-425-27	Sequence 27, Appli
32	128.2	10.9	600	12	US-10-255-433-27566	Sequence 27566, A
33	126.6	10.8	1320	15	US-10-369-493-46157	Sequence 46157, A
34	126.4	10.8	357	13	US-10-125-968-977	Sequence 977, App
35	124.6	10.6	1784	13	US-10-339-351-4	Sequence 4, Appli
36	123.8	10.6	893	9	US-09-770-445-492	Sequence 492, Appli
37	121.4	10.3	1784	13	US-10-339-351-6	Sequence 6, Appli
38	120.4	10.3	480	11	US-09-918-985-35234	Sequence 35234, A
39	119.4	10.2	1218	12	US-10-369-493-46157	Sequence 46157, A
40	117.4	10.0	636	9	US-09-925-297-278	Sequence 278, Appli
41	112.2	9.6	2107	15	US-10-206-619-2	Sequence 2, Appli
42	111.6	9.5	1041	13	US-10-324-131-10	Sequence 10, Appli
43	108.8	9.3	1420	13	US-09-971-392-43	Sequence 43, Appli
44	104.6	8.9	1096	13	US-10-322-746-1	Sequence 1, Appli
45	104.2	8.9	1119	13	US-10-322-746-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-893-525-6
Sequence 6, Application US/09893525
Publication No. US20030126631A1
GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
FILE REFERENCE: 9369-172
CURRENT APPLICATION NUMBER: US/09/893,525
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 09/210,843
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: US 08/846,021
PRIOR FILING DATE: 1997-04-25
PRIOR APPLICATION NUMBER: US 08/366,783
PRIOR FILING DATE: 1994-12-30
PRIOR APPLICATION NUMBER: US 08/142,418
PRIOR FILING DATE: 1993-11-16
PRIOR APPLICATION NUMBER: US 07/659,835
PRIOR FILING DATE: 1991-02-22
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 2733
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Oleosin-Chymosin Fusion
NAME/KEY: CDS
LOCATION: (850)..(1203)
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: (1444)..(2727)
OTHER INFORMATION:
US-09-893-525-6
Query Match 76.3%, Score 894.8, DB 11, Length 2733;

Db 748 ACAACATCTGGAACACGAGGCGCTGGTTTCTCAGAACCTTCTCTGTCTACCTCAGCGCG 689
Qy 680 ATGGCGAG--GAGAGCATGCTCAAGCTTGAGCATATGATCCATCCATACACAGAT 736
Db 688 ATGACCAAGTGGCGAGCGGTGTGATCTTTGTGGATGAGCTCTTCTTACTACACTGAA 629
Qy 737 CTCTTCACTGGGTTTCCAGTCACTGTCAGCAGTACTGSCAAATTCAGTGAAGTGTCA 796
Db 628 GTCTGAATGGGTGCTCTTACCGTCAGGGGTACTGCAATACCGTGAACAGATCA 569
Qy 797 CCATCAGCGGTGTGGTGTGTCATGTGAAGGTGATGTCAACTATCTTGATACCGGTA 856
Db 568 CCATGAACGAGAGGCGCATCGCTGCGCTGAGGGGTGCGACAGGCGATGTGTGACACGGCA 509
Qy 857 CGTCCAGCTGTGGAGCTGACGAGCGGACATTTCTCAATTCAGAACCTATGAGGCA 916
Db 508 CCTCTGTCTTACCGGCGCCCAACGAGCCCATTTGCCAATCCAGAGCGACATCGGAGCA 449
Qy 917 CACAGAACCAAGTACGAGTGTGATGATAGATTTGGGACACCTTAGCTACATGCTACAG 976
Db 448 GCGAGAACTCAGATGAGCGACATGTGTGAGTGTGCTGAGCCATCAGAGCTGCCCCGCA 389
Qy 977 TTGTCTTGAATCAACGCGCAAGATGTACCACTGACCCCTCCGCTTATACAGCCGAG 1036
Db 388 TCGTCTTACCATCAATGAGAGTCCAGTACCCCGTCCAGCCAGTGTCTTACATCTGAGCA 329
Qy 1037 ATCAAGGTTTCCGACCGTGTGATTCAGAGTGAAGAACCTATCC-----CAGA 1084
Db 328 GCGAGGGAGGTGCTCATGAGTGTCTTCCAGGGCATATACCTCCCAAGATCTGAGAGC 269
Qy 1085 AATGATCTTGGGAGATGTGTTCAATTGATGATACAGAGGCTTTGACAGGGCGACCA 1144
Db 268 TTGGATCTTGGGTATGTCTTCACTCCGCGAGTACTTACGCTTTCAGAGGGAACA 209
Qy 1145 ACCTCGTTGGGCTAGC 1160
Db 208 ACCAGGTGGGCTGSC 193

RESULT 3
US-10-074-978A-7/c
Sequence 7, Application US/10074978A
Publication No. US20040010119A1
GENERAL INFORMATION:
APPLICANT: Leite, Mario
APPLICANT: Spytek, Kimberly A
APPLICANT: Guo, Xiaojia (Saaha)
APPLICANT: Fernandes, Elma
APPLICANT: Li, Li
APPLICANT: Kekuda, Ramesh
APPLICANT: Liu, Xiahong
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Perenc
APPLICANT: Paturajan, Meera
APPLICANT: Ballinger, Robert
APPLICANT: Vermet, Corine
APPLICANT: Tchernev, Velizar T
APPLICANT: Malyankar, Uriel M
APPLICANT: Gusev, Vladimir
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter S
APPLICANT: Ellerman, Karen
APPLICANT: Heyes, Melvin P
APPLICANT: Herman, John
APPLICANT: Pena, Carol E A
APPLICANT: Shinkets, Richard A
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Moore, No. US20040010119A11le
APPLICANT: Shenoy, Suresh
APPLICANT: Edinger, Shlomlt
APPLICANT: Gunther, Erik

APPLICANT: Stone, Dave
APPLICANT: Miller, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glenda
FILE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-269
CURRENT APPLICATION NUMBER: US/10/074, 978A
PRIOR FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: 60/268, 221
PRIOR FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/335, 109
PRIOR FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: 60/312, 284
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/268, 496
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: 60/276, 703
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/330, 293
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/322, 127
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/280, 899
PRIOR FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: 60/310, 797
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/268, 646
PRIOR FILING DATE: 2001-02-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 547
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 1131
TYPE: DNA
ORGANISM: Homo sapiens
US-10-074-978A-7

Query Match 30.6%; Score 359.2; DB 12; Length 1131;
Best Local Similarity 60.3%; Pred. No. 1.1e-114;
Matches 661; Conservative 0; Mismatches 408; Indels 27; Gaps 3;

Qy 92 TTCTCTCTCAAAAGTAAAGTCTCTCGTAAAGCGCTGAAGAAACATGAACTTCTAGAG 151
Db 1112 TCCCCCTATCAGAAAGAGTCTTGAAGGCGCACCTGTCCGAGGTGGCTGTGAAG 1053
Qy 152 ACTTCTTGCAAAACAGATGAGCATGAGCAGAGCAAGTACTCCGGCTTCGGTGAATTG 211
Db 1052 ACTTCTTGAAAGAACACAACTCAACCCAGCCAGAAAGTACTTCCCGAGTGGAGGCTC 993
Qy 212 CTAGCGT-----GGCACTTACCACTACCTGTATGATGATCACTTGGAGAGA 259
Db 992 CCACCTGTGATGATGAACAGCCCTGAGAACTACCTGATATGAGATGAGTCTTGGACCTA 933
Qy 260 TCTACTCGGAACCCCGCTCAAGAGTTCAACGTTCTTGTGATACGTGCTCTGACT 319
Db 932 TCGGATCGAATCTCTGCGCAGATTTCAACCGTCTTGTGACACGGCTCTCTCAACC 873
Qy 320 TCTGGGTTCCCTTATCTACTGCAAGAGCAATGCTGCAAGAACCCAAAGATTGATC 379
Db 872 TGTGGGTGCCCTCAGTCTACTGCTCAGTCTTGTGCTGCAACCAACACCGCTTCAAC 813
Qy 380 CGAAGAACTGTCACCTTCCAGAACTTGAAGCAACCTTGTCTATACATGAGTACAG 439
Db 812 CTGAGATCTTTCACCTTACCAAGTCCACAGCGAGACAGTCTTCCATCCATCGGACCG 753
Qy 440 GTAGCATCAAGATCTTAGGCTATGATACCGTCACTGTCTCCAACTTGTGACATTC 499
Db 752 GCAAGATCAAGGATCTCGGATTCAGCACTGTCCAGTTGAGGAGATCTCTAGACCA 693
Qy 500 AACGACATGAGACTTGAACCCAGAGACAGAGTATGTTCACTATGACAGATTG 559
Db 692 ATCAGATCTTGGGCTGAGGAGAGAGAACTGCTCTTCTGTATATGCTCCCTTCG 633

QY 560 ATGGCATCTCTGGTATGCAATACCATCGCTCGGCTCAGAGTACTGATPACTGTGTGTTG 619
 Db 632 ATGGCAATCTGGGGCTGGCTTACCCAGCATTTCTCTCCGGGGCCACACCCGCTCTTG 573
 QY 620 ACAACATGATGAAACGACACCTAGTACTGATCAAGTCTGTCTGTGTTTACATGAGACGA 679
 Db 572 ACAACATCTGGAACGAGGCTGTGTTTCTCAGAACCTCTCTGTCTTACTGAGGCCG 513
 QY 680 ATGGCCAG--GAGAGATGCTCAACGCTTGAGCTATTGATCCATCTTACACAGGAT 736
 Db 512 ATGACCAAGATGGCAGCGTGTATCTTTGTGTGGCAATGACTCTTCTTACTACCTGAA 453
 QY 737 CTCTTACGAGGTTTCCAGTACTGTGAGAGATGAGTACGAGCAATTCCTGTGAGACGTCTA 796
 Db 452 GTCTGAACTGGGCTGCTTACCTGAGGGTTTACTGGCAGATCACCGTGAACAGATCA 393
 QY 797 CCATGACGAGTGTGTTGTTGTCATGTAAGTGAATGATCAAGCTATCTTGATACCGGTA 856
 Db 392 CCATGAAACGAGAGGCGCATGCTGCGCTGAGGGCTGCCAGGCTTGTGACACCGGCA 333
 QY 857 CGTCAAGCTGTGCGGACCTAGACGACCAATTTCTCAATTCAGCAAGCTATTGAGGCA 916
 Db 332 CCTCTGTGACCGGCGCCCAACGAGCCCATTCGCAACATCCAGAGCGCATGCGAGCCA 273
 QY 917 CAGAGAACGATAGGAGTGTGATGACATAGTTCGAGCAACCTTAGTACATGCTTACAG 976
 Db 272 GCGAGAACTCAGATGGGAGATGTGTGTGCTGCTCAGCATCAGAGGCTTCCCGACA 213
 QY 977 TTGTCTTGGATCAACGCGCAAGATGATGACCCCTCGCTTATACAGGACGAG 1036
 Db 212 TGTCTTTCACCATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 153
 QY 1037 ATCAAGGTTTCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATG 1084
 Db 152 GCGAGGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 93
 QY 1085 AATGATCTTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1144
 Db 92 TTGTGATCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 33
 QY 1145 ACCTCGTGGGCTAGC 1160
 Db 32 ACCAGGTGAGCTTGGC 17

RESULT 4
 US-10-074-978A-3
 ; Sequence 3, Application US/10074978A
 ; Publication No. US20040010119A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Leite, Mario
 ; APPLICANT: Spytek, Kimberly A
 ; APPLICANT: Guo, Xiaojia (Sasha)
 ; APPLICANT: Fernandes, Elma
 ; APPLICANT: Li, Li
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Boldog, Ferenc
 ; APPLICANT: Patlurajan, Meera
 ; APPLICANT: Bialock, Angela
 ; APPLICANT: Ballinger, Robert
 ; APPLICANT: Verneil, Corine
 ; APPLICANT: Tchernev, Velizar T
 ; APPLICANT: Malyankar, Uriel M
 ; APPLICANT: Gusev, Vladimir
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Mezes, Peter S
 ; APPLICANT: Elleman, Karen
 ; APPLICANT: Heyes, Melvin P
 ; APPLICANT: Herrman, John
 ; APPLICANT: Pena, Carol E A
 ; APPLICANT: Shimkets, Richard A

APPLICANT: Taupier Jr, Raymond J
 APPLICANT: Moore, No. US20040010119A111e
 APPLICANT: Shenoy, Suresh
 APPLICANT: Edinger, Shlomit
 APPLICANT: Gunther, Erik
 APPLICANT: Stone, Dave
 APPLICANT: Millet, Isabelle
 APPLICANT: Peyman, John
 APPLICANT: Smithson, Glenda
 TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
 FILE REFERENCE: 21402-269
 CURRENT APPLICATION NUMBER: US/10/074, 978A
 CURRENT FILING DATE: 2003-01-07
 PRIOR APPLICATION NUMBER: 60/268, 221
 PRIOR FILING DATE: 2001-02-12
 PRIOR APPLICATION NUMBER: 60/335, 109
 PRIOR FILING DATE: 2001-10-31
 PRIOR APPLICATION NUMBER: 60/312, 284
 PRIOR FILING DATE: 2001-08-14
 PRIOR APPLICATION NUMBER: 60/268, 496
 PRIOR FILING DATE: 2001-02-13
 PRIOR APPLICATION NUMBER: 60/276, 703
 PRIOR FILING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: 60/330, 293
 PRIOR FILING DATE: 2001-10-18
 PRIOR APPLICATION NUMBER: 60/322, 127
 PRIOR FILING DATE: 2001-11-21
 PRIOR APPLICATION NUMBER: 60/280, 899
 PRIOR FILING DATE: 2001-04-02
 PRIOR APPLICATION NUMBER: 60/310, 797
 PRIOR FILING DATE: 2001-08-08
 PRIOR APPLICATION NUMBER: 60/268, 646
 PRIOR FILING DATE: 2001-02-14
 REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 547
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 3
 LENGTH: 1688
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-074-978A-3

Query Match 30.6%; Score 359.2; DB 12; Length 1688;
 Best Local Similarity 60.3%; Pred. No. 1.4e-114;
 Matches 661; Conservative 0; Mismatches 408; Indels 27; Gaps 3;
 QY 92 TTCTCTCTACAAAGTAAAGTCTCTCTGTAAGGCGCTGAAAGAAATGAGACTTCTGAAG 151
 Db 412 TCCCTCTCTACAAAGAAAGTCTCTGAGGCGCACCTGTCCAGAGCTGCTGTAAGG 471
 QY 152 ACTTCTTGCAAAACAAAGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 211
 Db 472 ACTTCTTGAAAGAAAGCAACCTCAACCCAGGAAAGTATCTTCCCGATGGAAGCTC 531
 QY 212 CTAGCGT-----GCCACTTACCACTACTCTTATATGTAATCTTGGGAAAG 259
 Db 532 CCACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 591
 QY 260 TCTACCTGGAACCCGCGCTCAAGATGATGATGATGATGATGATGATGATGATGATG 319
 Db 592 TCGGCACTGGAATCTCTGCGCAAGATTTCACTGTCTTTTGAACCGGCTCTCAAC 651
 QY 320 TCTGAGTCTCTCTATCTACTGCAAGAGATGCTGCAAGAACCAACCAAGATTCGATC 379
 Db 652 TGTGGGTGCTCTAGCTACTGTCTGCAAGTCTTGTGCAACCAACCAACCGCTTCAAC 711
 QY 380 CAGAAAGTGTCTCACTTCTGCAAGTATGAGCAACCTTGTCTATACATGAGTACAG 439
 Db 712 CTGAGATCTCTTCACTTCAAGGCGCACCAAGCAAGTCTTCACTTCACTTCAAGGCG 771
 QY 440 GTAGCATGCAAGAACTTATGAGTATGATGATGATGATGATGATGATGATGATGATG 499
 Db 772 GAGCATGACAGGATCTCTGAGTATGAGCATGTCTCAGGTTGAGGCAATCTTGAACCA 831

```

QY 500 AACAGACAGTAGACCTAGACCCAGAGACAGTAGATGCTTCACTTAGAGAAATTG 559
DB 832 ATCAGATCTTGGCCTGAGCGAGAGAGAACTGGCTCTCTCTGATATATGCTCCCTTC 891
QY 560 ATGAGATCTTGGATGATGAGATACCCATGCTGCGCTGAGATGATGATGATGATGATG 619
DB 892 ATGAGATCTTGGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 951
QY 620 ACAATATGATGATGAGACAGACCTAGATGATGATGATGATGATGATGATGATGATGATG 679
DB 952 ACAATATGATGATGAGACAGACCTAGATGATGATGATGATGATGATGATGATGATGATG 1011
QY 680 ATGAGACAG--GAGAGACATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 736
DB 1012 ATGAGACAGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1071
QY 737 CTCTTCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 796
DB 1072 GTCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1131
QY 797 CCATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 856
DB 1132 CCATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1191
QY 857 CGTCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 916
DB 1192 CCTCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1251
QY 917 CACAGACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 976
DB 1252 GCGAGACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1311
QY 977 TTGCTTGAATGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1036
DB 1312 TCGCTTGAATGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1371
QY 1037 ATCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1084
DB 1372 GCGAGACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1431
QY 1085 AATGATCTTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1144
DB 1432 TTGATCTTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1491
QY 1145 ACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1160
DB 1492 ACCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1507

```

RESULT 5
US-10-074-978A-5/c

```

; Sequence 5, Application US/10074978A
; Publication No. US20040010119A1
; GENERAL INFORMATION:
; APPLICANT: Leite, Mario
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Fernandes, Elma
; APPLICANT: Li, Li
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Liu, Xiaohong
; APPLICANT: Cauman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Patucurajan, Meera
; APPLICANT: Blalock, Angela
; APPLICANT: Baillinger, Robert
; APPLICANT: Vernet, Corine
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Gusev, Vladimir
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter S

```

```

; APPLICANT: Ellerman, Karen
; APPLICANT: Heyes, Melvin P
; APPLICANT: Herrman, John
; APPLICANT: Pena, Carol E A
; APPLICANT: Shinkets, Richard A
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Moore, No. US20040010119A11e
; APPLICANT: Shenoy, Suresh
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, Dave
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-269
; CURRENT APPLICATION NUMBER: US/10/074,978A
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: 60/268,221
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/335,109
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/312,284
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/268,496
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 60/276,703
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/330,293
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/322,127
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/280,899
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/310,797
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/268,646
; PRIOR FILING DATE: 2001-02-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-074-978A-5

```

Query Match 30.5%; Score 357.6; DB 12; Length 1131;
Best Local Similarity 60.2%; Pred. No. 4,1e-114;
Matches 660; Conservative 0; Mismatches 409; Indels 27; Gaps 3;

```

QY 92 TTCCTCTCAAGAGTAAGTCTCTCCGTAAGCGCTGAGAGAACATGAGACTTCTAGAG 151
DB 1112 TCCCTCATCAGAAAGAGTCTTGAAGCGCGACCTCTCGAGAGCTGCTGAGAG 1053
QY 152 ACTTCTGAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 211
DB 1052 ACTTCTGAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 993
QY 212 CTAGAGT-----GCCACTTACCACTACTGATGATGATGATGATGATGATGATGATG 259
DB 992 CCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 933
QY 260 TCTACTGAGAACCCCGCTCAAGAGTTCACCGTCTCTTGTATGATGATGATGATGATGATG 319
DB 932 TCGGATCGAGAACTCTGCGCAGATTTCACTGCTCTTGTAGACACCGGCTCTCAAC 873
QY 320 TCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 379
DB 872 TGTGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 813
QY 380 CGAGAAAGCTGCTGCACTTCCAGAACTTAGGCAACCTTGTCTATACCTTAGGCTAGAG 439

```

```

Db      812 CTGAGAGATTCTTCCACTACCACTGACCAAGAGAGAGAGTCTTCATCCTACCTAGGACCG 753
Qy      440 GTAGCATGCAAGAAATCTTAGGCTATGANTACCGTCACTGTCTCCAACTTTGGACATTC 499
Db      752 GCAGCATGACAGGATCTCTCGATACGACACTGTCCAGGTTGAGGACATCTTGACACCA 693
Qy      500 AACAGACATGAGATCTTAGACACCAAGAACAGATGATGTCTTCACTATGCAAGATTGG 559
Db      692 ATGAGATCTTCCGGCTGAGGAGAGAGAGACCTGCTCTTCCTGTATTAATGCTCCCTTCG 633
Qy      560 ATGGCATCTCTGGTATGATGATACCACTGCTCGCTGAGATGACTGATATCTGTGTTG 619
Db      632 ATGGCATCTCTGGGCTGCTGCTTACCCAGCATTTCTCTCCGGGGCCACACCCGTCTTTG 573
Qy      620 ACAACATGATGAACCGACACCTAGTAGTCTCAAGCTTTGTTCTGCTTTTACATGAGACGA 679
Db      572 ACAACATCTGGAACCAAGGGGCTG6TGTCTCAGAGACCTCTTCTGTCTACCTCAGGCGG 513
Qy      680 ATGGCCAG---GAGAGATGCTCAGCGTTTGGAGCTATTGATCCATCTCTACTACACAGAT 736
Db      512 ATGACCAAGATGAGAGGAGTGTGATCTTTGTGTGCAATGACTCTTCTTACTACACCTGGA 453
Qy      737 CTCTTCACTGGGTTCCAGTCACTGTGAGAGAGTACTGGCAATTCATGAGACAGTGTCA 796
Db      452 GTCTGAATCTGGGTGCTGTACCGTCAAGGTTACTGGCAGATCACCTGTGAACAGATCA 393
Qy      797 CCATCAGCGGTGTGTGTGTCATGTGAAAGTGTGATGCAAGTATCTTGATACCGGTA 856
Db      392 CCATGAACGAGAGAGGCGATCGCTCGCTGAGGAGGCTCCAGGCGCATTTGTGACACGGGCA 333
Qy      857 CGTCCAAAGTGTGTGCGACCTTAGACAGGACATTCCTCAACATTTGACAAAGCTATTGGAGCCA 916
Db      332 CCTCTGTGCTGACCGGCGCCCAACAGCCCATGTGCAACATTCAGAGGACATCGGAGCCA 273
Qy      917 CACAGAACCAAGTACGAGTGAATTTGACATTAATTTGCGAACACTTAGCTATCATGCTTACAG 976
Db      272 GCGAGAACTCAGATGAGGAGATGTGTGTCACTGTCTCAGCATCAGAGAGCTGTCCGACA 213
Qy      977 TTGTCTTTGAGATCAACGAGAAATGTATACCACTGACCCCTCCGCTTATACGAGCCAGG 1036
Db      212 TCGTCTTCAACATCAATGAGATGATCCAGTACCCCGTGCACCAAGTGCCTTATCCTGAGA 153
Qy      1037 ATCAAGGTTCTGACCAAGTGTGATTTCCAGAGTGAAGCAATTC-----CAGA 1084
Db      152 GCGAGGAGGAGCTGATCAGTGTGCTTCCAGGCGATGAACCTCCCAACGAACTGTGAGAGC 93
Qy      1085 AATGATCTTGGGAGATGTGTTCATTCGTGTGATCTACAGGCTTTGACAGGGGCCAACA 1144
Db      92 TTGGAATCTCTGGGTGATGTCTTTCATCCGCAATCTTTACCGTCTTGAACAGGCAACCA 33
Qy      1145 ACCCTCGTTGGGCTAGC 1160
Db      32 ACCAGGTCAAGCTGCG 17

```

RESULT 6
US-10-074-978A-11/c

Sequence 11, Application US/10074978A
Publication No. US20040010119A1

GENERAL INFORMATION:

APPLICANT: Velite, Mario
APPLICANT: Spytek, Kimberly A
APPLICANT: Guo, Xiaojia (Sasha)
APPLICANT: Fernandez, Elma
APPLICANT: Li, Li
APPLICANT: Ketuda, Ramesh
APPLICANT: Liu, Xiaohong
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Patuturajan, Meera
APPLICANT: Blalock, Angela
APPLICANT: Ballinger, Robert
APPLICANT: Vermet, Corine

```

APPLICANT: Tchernev, Velizar T
APPLICANT: Malyankar, Uriel M
APPLICANT: Gusev, Vladimir
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter S
APPLICANT: Ellerman, Karen
APPLICANT: Heyes, Melvin P
APPLICANT: Herrman, John
APPLICANT: Pena, Carol E A
APPLICANT: Shimketa, Richard A
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Moore, No. US20040010119A11le
APPLICANT: Shenoy, Suresh
APPLICANT: Edinger, Shlomit
APPLICANT: Gunther, Erik
APPLICANT: Stone, Dave
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glenda
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-269
CURRENT APPLICATION NUMBER: US/10/074,978A
CURRENT FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: 60/268,221
PRIOR FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/335,109
PRIOR FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: 60/312,284
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/268,496
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: 60/276,703
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/330,293
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/322,127
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/280,899
PRIOR FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: 60/310,797
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/268,646
PRIOR FILING DATE: 2001-02-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 547
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 1131
TYPE: DNA
ORGANISM: Homo sapiens
US-10-074-978A-11

```

Query Match 30.5%; Score 357.6; DB 12; Length 1131;
Best Local Similarity 60.2%; Pred. No. 4.1e-114;
Matches 660; Conservative 0; Mismatches 409; Indels 27; Gaps 3;

```

Qy      92 TTCTCTCTCAAAAGGTAAGTCTCTCCGTAAGCGCTGAGAAACAATGACTTCTAGAAG 151
Db      1112 TCCCTCTCATGAAAGAAAGTCTTGTAGGCGGACCTGTCCAGAGCTGCTGTGAAG 1053
Qy      152 ACTTCTTGCAAAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 211
Db      1052 ACTTCTTGAAAGCAACCTTCACCCAGGAGGAGAAAGTATCTTCCAGTGGAGGCTC 993
Qy      212 CTAGCGT-----GCCACTTACCACTACTGATGATGATGATGATGATGATGATGATG 259
Db      992 CCACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 933
Qy      260 TCTACCTGGAACCCGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 319
Db      932 TCGGCATCGGAATCTCTGCGCAGAGATTTTCAAGTGTCTTTTGAACAGGCTCTTCAACC 873
Qy      320 TCTGGGTCTCTATCTACTGCAAGAGCAATGCTTGAAGAACCAACAAAGATTGATC 379

```



```

Db      872 TGGGGTGGCCCTCAGTCTAATGCTCCAGTCTTGCCCTGCAACCAACCCCTTCAACC 813
Qy      380 CGAGAAAGTGGTCCACCTTCCAGAACTTAGGCAAACTTGTCTATACGTAAGGACG 439
Db      812 CTGAGGATTTCTTCCACTTACAGTCCACGAGAGACAGTCTTCCACTACCTACGAGCCG 753
Qy      440 GTAGCATGCAAGAACTTATAGCTATGATACCGTCACTGTCTCAACATTGTGACATTTC 499
Db      752 GGAGCATGACAGGCAATCCCTGGATTCAGACAGTCCAGGTTGGAGGCACTCTGACACCA 693
Qy      500 AACAGACAGTACGATCTTACGACCCAGAAACAGAGTATGTCTTCACTATGCAAAATTGG 559
Db      692 ATCAGATCTTCCGCTGAGCGAGACGGAACCTGGCTCTCTCTGATATATGCTCCCTTGG 633
Qy      560 ATGGACCTCTTGTATGATATACCATTCCTGCGCGGCAAGATCTGATACCTGTGTTG 619
Db      632 ATGACATCTGGGGCTGGGCTTACCCAGCATTTCTCTCCGGGGCCACACCCGCTTGG 573
Qy      620 ACAACATGATGAAACGACACCTAGTACGTAAGATCTTGTCTGCTTACATGACAGGA 679
Db      572 ACAACATCTGAAACGAGGCTGTTTCTCAGGACCTCTTCTGTCTACCTCAGGCGG 513
Qy      680 ATGGCCAG--GAGACATGCTCAAGCTTGAAGTATGATCCATCTAACAAGAT 736
Db      512 ATGACACAGATGAGGACGCTGATCTTGTGGGCAATGACTCTTCTTACTACACTGGA 453
Qy      737 CTCTTCACTGGGTTCCAGTCACTGTGACAGACGTAAGTCACTGTGACAGTGTGA 796
Db      452 GCTTAACTGGGTTGCTGTACCGTCAAGGAGTTCAGGAGATCAACCGTGAACAGATCA 393
Qy      797 CCATCAGCGGCTGATGTTGTGATGAAAGTGAATGATCACTATCTTGAATCCGGA 856
Db      392 CCATGAGGAGAGAGGCACTGCTGCTGAGGCTGCGAGGCTCATTTGTGACCCGGA 333
Qy      857 CGTCAAGCTGTGCGAAGCTTACAGAGCAATTTCTCAACATTGAGCAAGCTTATGAGCA 916
Db      332 CCTCTGTGTCAGCGGCGCCCAACAGGCCATTGCCAATCAAGAGCGCATCGAGGCA 273
Qy      917 CACAGAAACGATGACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 976
Db      272 GCGAAGAACTCAGATGACGATGATGATGATGATGATGATGATGATGATGATGATG 213
Qy      977 TTGTCTTGAAGATCAACGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1036
Db      212 TCGTCTTCACTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 153
Qy      1037 ATCAAGGCTTCTGCAACAGTGTGATGATGATGATGATGATGATGATGATGATGATG 1084
Db      152 GCGAGGGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 93
Qy      1085 AATGATCTTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1144
Db      92 TTTGATCTCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 33
Qy      1145 ACCTGCTGGGCTAGC 1160
Db      32 ACCAGGTACGCTGCG 17

```

```

APPLICANT: Boldog, Ferenc
APPLICANT: Paturajan, Meera
APPLICANT: Ballock, Angela
APPLICANT: Ballinger, Robert
APPLICANT: Verne, Corine
APPLICANT: Tchenev, Veilzar T
APPLICANT: Malyankar, Uriel M
APPLICANT: Gusev, Vladimir
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter S
APPLICANT: Elleman, Karen
APPLICANT: Heyes, Melvin P
APPLICANT: Herman, John
APPLICANT: Pena, Carol E A
APPLICANT: Shinkets, Richard A
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Moore, No. US2004001019A111e
APPLICANT: Shenoy, Suresh
APPLICANT: Edinger, Shlomit
APPLICANT: Gunther, Erik
APPLICANT: Stone, Dave
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glenda
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-269
CURRENT APPLICATION NUMBER: US/10/074,978A
CURRENT FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: 60/268,221
PRIOR FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/335,109
PRIOR FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: 60/312,284
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/268,496
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: 60/276,703
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/330,293
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/322,127
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/280,899
PRIOR FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: 60/310,797
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/268,646
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 547
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 1131
TYPE: DNA
ORGANISM: Homo sapiens
US-10-074-978A-9
Query Match 30.3%; Score 356; DB 12; Length 1131;
Best Local Similarity 60.1%; Pred. No. 1.5e-113;
Matches 659; Conservative 0; Mismatches 410; Indels 27; Gaps 3;

```

```

RESULT 7
US-10-074-978A-9/c
Sequence 9, Application US/10074978A
Publication No. US20040010119A1
GENERAL INFORMATION:
APPLICANT: Leite, Mario
APPLICANT: Spytek, Kimberly A
APPLICANT: Guo, Xiaojia (Sasha)
APPLICANT: Fernandes, Elma
APPLICANT: Li, Li
APPLICANT: Kerkuda, Ramesh
APPLICANT: Liu, Xiaohong
APPLICANT: Casman, Stacie

```

```

Qy      92 TTCTCTTCAAAAGCTATGCTCTCCGTAAGGCGCTGAAGAAATGATGATGATGATG 151
Db      1112 TCCCTCTATCAGAAAGAGTCTTGAAGGCGCACCTGTCCAGAGTGGCTGTGAAGG 1053
Qy      152 ACTCTTSCAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 211
Db      1052 ACTTCTGTAAGAGAGCAACCTCAACAGGCAAGAAATGATGATGATGATGATGATG 993
Qy      212 CTAGGCT-----GCCATTACCAACTACCTTGTATGATGATGATGATGATGATG 259
Db      992 CCACCCTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 933

```


501 ACAGACAGTAGAGCTTAGACCCAGAACACAGGTGATGTTCTTCACTATGACAAATTGCA 560
 535 CCAGCAGCTTTGGAGAAAGTGTACAGAGCCAGAGCCAGACTTTGTGTGATGACAGATTGA 594
 561 TGGCATCTCTTGTGATGATACCCATGCTGCGCTGACAGTACTGATACCTGTGTTGA 620
 595 TGGAAATCTGGGCTCGGGATACCCCTCTTGGCTGTGGAGAGTGAATCCAGTATTGTA 654
 621 CAACATGATGAACCGACACCTAGTAGCTCAAGACTTCTCGGTTTATACAGACAGAA 680
 655 CAACATGATGCTGACAACTGCTGATGCTGCGCATGTTTCTGTCTATACAGACATGA 714
 681 -----TGGCAGAGAGCATGCTCAACGCTTGGAGCTATGATTCATCTTACTACAC 731
 715 CCAGAAAGTGTGGGGAGGAGAGAGCTGATTTTGGAGGCTAGACCACTCCACTTTCTC 774
 732 AGGATCTCTTCACTGGGTTCCAGTCACTGTGACAGATGCTGCAATTCATCTGTGACAG 791
 775 TGGGAGCTGAAATGGGTTCCAGTCAACCAAGCAAGCTTACTGCGAGATTGCACTGGATTA 834
 792 TGTCAACATCAGCGGTGTGTTGTGATGAGTGAAGTGTGATGATCTTGTGATAC 851
 835 CATCAAGGTGGAGAGCACTGTATGTTCTGCTCCAGGGCTGCGAGCCATTTGTGACAC 894
 852 CGGTACGTCCAGCTGTGTGCACTAGCAGCAGCAATTCTCAACTTCAAGCACTATTGG 911
 895 AGGACTTCCCTCATCACTGCGCTTCCGACAAAGATTAGCAGCTGCAAAAGCCATTGG 954
 912 AGCCACACAGAAACAGTACGCTGAGTTTGAATGATGCGACAACTTGTGCTTACAGTGC 971
 955 GGCAGGCC---CCGTGATGAGAAATATGCTGTGAGTGTGCAACTTAACTGATGCGC 1011
 972 TACAGTGTCTTGTGATGAGCAAGATGTAACCACTGACCCCTCCGCTATACAG 1031
 1012 GATGTCACTTCACTTCACTTAAAGAGTCCCTTATACCTTCAAGCCAACTGCTACACCT 1071
 1032 CCAGATCAAG-----GGTCTGACACAGTGTGATTCAGAGTGAACCATTC 1079
 1072 ACTGATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
 1080 CCAGAA-----TGTGATCTTGGAGATGATGATGATGATGATGATGATGATGATGAT 1127
 1132 CCCTCAGCTGGGCCCCCTGTGATCTGTGGGATGCTTCAATTCAGACGTTTACTCAGT 1191
 1128 CTTTGAACAGGCGCAACACTGTTGGGCTTACTTAAGCAATC 1170
 1192 CTTTGAACGTTGGAAATTAACCGTGTGGACTGGCCCCAGCAGTC 1234

RESULT 9
 US-10-101-510-457
 ; Sequence 457, Application US/10101510
 ; Publication No. US20030148295A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WAN, JACKSON
 ; APPLICANT: WAN, YIXIN
 ; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
 ; FILE REFERENCE: 15117.0012
 ; CURRENT APPLICATION NUMBER: US/10/101,510
 ; PRIOR FILING DATE: 2002-03-20
 ; PRIOR APPLICATION NUMBER: 60/276,947
 ; NUMBER OF SEQ ID NOS: 805
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 457
 ; LENGTH: 2201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: modified_base
 ; LOCATION: (1432)..(1467)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-101-510-457

Query Match 21.6%; Score 253; DB 13; Length 2201;
 Best Local Similarity 55.3%; Pred. No. 3.6e-77;
 Matches 621; Conservative 0; Mismatches 460; Indels 42; Gaps 5;
 87 CCGATTCCTCTTACAAAGGTAAGTCTCTCGTAAGGCGCTGAAGAACTGACTTCT 146
 158 CAGGTGCCCCCTCAGAGAGCATCCGCTCCCAAGAAAGAGTGGCGGACCGAGCCACT 217
 147 AGAAGCTTCTTGCAGAAACAAAGTAGGATCA-----GCAGAGTACTCGGCTT 200
 218 CTGAGTCTTGGAAATCCCAATATTGGACATGATCCAGTTCAACCGAGTCTGTCAAT 277
 201 CGGTAAAGTGTGAGCGTCCAGCACTTACCACTTATGATGATCAATCTTTGGAGAAAT 260
 278 GAGCAGAGTGCAGAGAAACCCCTCATACACTTGTGATATGAAATCTTCCGACATAT 337
 261 CTACCTCGAAACCCCGCTCAAGAGTTCACCGTTCTCTTGAATACCTGTTCTCTGACTT 320
 338 CTGATGCTCCCGACCAAGAACTTCACTGTATCTTGCATCTGCTCCCTCCACCT 397
 321 CTGGGTTCCCTCTATCTCTGAGAGCAATGCTGCAAGAACCAAGATTGATTC 380
 398 CTGGGTCCCTCTGTGTCTGATCACTGACCCAGCCGCAAGACGCAACAGGTTCCAGCC 457
 381 GAGAAAGTGTCCACCTTCCAGAACTTAGGCAACCTTGTCTATACCTACGTTACAG 440
 458 TTCCAGTCCAGCAATACAGCCAGCCAGTCAATCTTCTCATCTTCAATGAAACGG 517
 441 TAGCATGCAAGAAATCTTAGGCTATGATACCGTCACTGTCTTCAATTTGAGACATTA 500
 518 GAGCTTGTCCGGAAATCATTTGAGGCGGACCAAGTCTCTGTGAGAGACATTAACGTTGG 577
 501 ACAGACATGAGACTTACAGACCCAAAGAACAGGTGATGTTCTTACCTTATGAGAAATTGA 560
 578 CCAGAGTTTGGAGAAAGTGTACAGAGCCAGGCGCAGCTTTGTGATGAGAGATTGTA 637
 561 TGGCATCTTGTATGATATCCATGCTGCGCTGCGCAAGATGATGATGATGATGATGATGAT 620
 638 TGGAAATCTGGGCTGGGATACCCCTCTTGTGCTGTGGAGAGTGAATCCAGTATTGTA 697
 621 CAACATGATGAACCGACACTAGTAGCTCAAGACTTGTCTCGTTTATGATGACAGAA 680
 698 CAACATGATGCTCAGAACTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 757
 681 -----TGGCAGAGAGCATGCTCAACGCTTGGAGCTATGATCCATCTTACTACAC 731
 758 CCAGAAAGTGTGTGGGAGGAGGAGGAGCTGATTTTGGAGGCTAACGACCACTCCATTTCTC 817
 732 AGGATCTTCACTGAGGTTCCAGTACCTGTGACAGATGATGAGCAATTCATCTGTGACAG 791
 818 TGGAGCTTGAATTTGGGTTCCAGTCAACCAAGCTTATCGGAGATTTGACAGATTA 877
 792 TGTCAACATCAGCGGTGTGTTGTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 851
 878 CATCCAGGTGGAGGACATGTTATGTTCTGCTCCAGAGGCTGCGAGGCAATTTGTGACAC 937
 852 CGGTACGTCAAGCTGTGCGACCTAGCAGAGCAATTCATCAACTTCAAGCACTATTGG 911
 938 AGGACTTCCCTCATCACTGCGCTTCCGACAAAGTTAAGAGGCTGCAAAAGCCATTGG 997
 912 AGCCACACAGAAACAGTACGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 971
 998 GGCAGGCC---CCGTGATGAGAAATATGCTGTGAGTGTGCAACCTTAACTGATATGCC 1054
 972 TACAGTGTCTTGTGATGATCAACGCAAGATGTAACCACTGACCCCTCCGCTATACAG 1031
 1055 GATGTCACTTCACTTCACTTAAAGAGTCCCTATACCTTCAAGCCCAATGCTTACACCTT 1114
 1032 CCAGATCAAG-----GGTCTGACACAGTGTGATTCAGAGTGAAGAACATTC 1079
 1115 ACTGACTTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1174

```
QY 1080 CCAGAAA-----TGGATCTTGGAGAGATGCTTCATTCGTGAGTACTACAGCCT 1127
Db 1175 CCTCCAGCTGGGCCCCCTCTGGATCTCGGAGAGTCTTCATTCAGACGATTTTACGAGT 1234
QY 1128 CTTTGAAGGGCCACACACCTCGTGGGCTAGCTAAAGCAATC 1170
Db 1235 CTTTGAACCGTGGGAATTACCGTGTGGGACTGGCCCCAGCAGTC 1277

RESULT 10
US-09-969-708-441
; Sequence 441, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 441
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-441

Query Match 21.2%; Score 248.4; DB 10; Length 1366;
Best Local Similarity 53.8%; Pred. No. 1.1e-75;
Matches 594; Conservative 0; Mismatches 486; Indels 24; Gaps 3;

QY 92 TTCCTCTCTCAAAAGTAAGTCTCTCCGTAGGCGCTGAAAGAAACATGACCTTCTAGAG 151
Db 1112 TGGCCCTGAAAGAAATTAAAGTCTATCCGTAGACCATGAAAGAGAGAGGGCTTGGTGGGG 171
QY 152 ACTTCTTGAGAAACACACAGTATGCGATCAGACGACGATCTCCGGCTTGGTGAAGTTG 211
Db 172 AGTTCTCGAGAGACCCACAAAGTATATCTGCTTGAAGTACCGCTTTGTGACCTCAGCG 231
QY 212 CTAGCGTGCACCTTACCACTACCTTATAGTCAATCTTTGGAGAATCTACCTCGGAA 271
Db 232 TGAACCTACGAGCCCATGCGCTACATGATGCTGCTACTTTGGAGATCAGCATCGGGA 291
QY 272 CCCCGCTCAAGAGTTGACCGTCTCTTTGATATGCTGTTCTCTGACTTCTGGTTCCT 331
Db 292 CTCACACCCCAAGACTCTGCTGCTTTTGAACCGGCTCCTCCAACTTGTGGGTGCGCT 351
QY 332 CTATCTCTCTGCAAGAGCAATGCTCTGCAAGAACCAACCAAGATTGATCCGAGAAAGTCT 391
Db 352 CTGTCTACTGCGCAGAGCGCCGTCACAGTCACTCCGCTTCAACCCAGCGAGTCTGT 411
QY 392 CCAACTTCCAGAACTTATGAGCAAAACCTTGTCTATATACCTACGCTAGAGATGAGCAAG 451
Db 412 CCACTCTACTCCACCAATGGGAGACCTTCTCCCTGAGTATGAGAGGACGCTCAGCG 471
QY 452 GAATCTTAGGCTATGATACCGTCACTGTCTCCAAATTTGAGACATTCACAGACAGTAG 511
Db 472 GCTTCTTTGGCTATGACACCTGACTGTCCAGAGCATTCAGAGTCCCAACAGAGAGTTG 531
QY 512 GACTTAGACACCCAGAAACAGAGTATGTCTTACACTTATGCAAAATTTGATGGCATCTTG 571
Db 532 GCTTGAATGAGAAAGACCTGTGTACCAACTTGTCTATGGCGAGTTTGAATGCGCATATGG 591
QY 572 GTATGGCATCCCATGCTGCGCGTCAAGATCTCGATACCTGTGTGTAACAAATGATGA 631
Db 592 GCCTGGCTTACCTGCTGCTGTGCTGTGATGAGGCCACCAACGCTATGAGGAGGCTGCTG 651
```

```
QY 632 ACCGACACCTAGTACTCAAGACTTGTCTCGTTTACATGACACAGAAATGCCACAGAGA 691
Db 652 AGAGAGGCGCCCTCAACACAGACCCCGTCTTCAGGGTCACTCAACACAGAGGGCTTCA 711
QY 692 GCAATGCTACGCG-----TTGAGAGTATGATTCATCTCCTACTACACAGGATCTTCTACT 745
Db 712 GCGGGGAGCGGTTGTTCTTTGGGGGTGTGATAGACGCTGTATACAGGGGCGAGTTACT 771
QY 746 GGGTTCCAGTCACTGTGACAGTACTGCAATTCATCTGTGACAGTGTCAACATCAGCG 805
Db 772 GGGCGCTGTTCACCCAGAACTCTACTGGCAGATTGGCATTTGAAGATTCCTCATGCGGG 831
QY 806 GTGTGG---TTGTGATGTGAAGTGTGATGTCAAGCTATCTTGAATACCGGTACTCCA 862
Db 832 GCCAGGCTTCGCGCTGTGTCTGTGAGGGTGTGCAGGCCATCTGTGACACAGGACACTTTC 891
QY 863 AGCTGTGCGACCTTACGACGACGACATTCCTCAACATTCAGACGATATTGAGGACACAGA 922
Db 892 TGCTAACTGTGCCCCAGACGATACATGATGTCTTCTTGAGGCGCACAGGGGCCAGAGAG 951
QY 923 ACCAGTACGGTGAAGTTTGACATATGATGCGACAACCTTATGCTACATGCTTACAGTTGTCT 982
Db 952 ATGAGTATGAGACAGTTTCTGTGAACTGTATGACGATTCAGAAATCTGCCAGCTTACCT 1011
QY 983 TTGAGATCAACGGCAAGATGTACCACTGACCCCTCGGCTTATACAGCCAGGATCAAG 1042
Db 1012 TCATCATCAATGATGTGTGAGGTTCCCTCTCCACCTTCTCTATATCTCAATACAAAG 1071
QY 1043 GGTTCGCAACGATGATTCGCAAGTGAACACATTC-----CAGAAAT 1087
Db 1072 GCTACTCAACCGTGGAGTGAAGCCCACTTACCTGTCTCCAGAAAGGCGACGCTGTCT 1131
QY 1088 GGAATCTTGGAGATGTGTTCATCTGTGATCTACTACAGCCTCTTTGACAGGGCCAAACAC 1147
Db 1132 GGAATCTCGGGAGATGTCTTCTCAGAGTCTTACATTCGCTTACAGACTTGGGCAACAA 1191
QY 1148 TCGTTGGGCTTACGTTAAAGCAATCT 1171
Db 1192 GAGTAGGCTTTTGCACCTGCGGCT 1215

RESULT 11
US-09-470-954A-45
; Sequence 45, Application US/09470954A
; Patent No. US20020169304A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: ASPARTYL PROTEASES
; FILE REFERENCE: PC-0008 US
; CURRENT APPLICATION NUMBER: US/09/470,954A
; CURRENT FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PERL Program
; SEQ ID NO 45
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g16s8285
US-09-470-954A-45

Query Match 21.2%; Score 248.4; DB 10; Length 1366;
Best Local Similarity 53.8%; Pred. No. 1.1e-75;
Matches 594; Conservative 0; Mismatches 486; Indels 24; Gaps 3;

QY 92 TTCCTCTCTCAAAAGTAAGTCTCTCCGTAGGCGCTGAAAGAAACATGACCTTCTAGAG 151
Db 1112 TGGCCCTGAAAGAAATTAAAGTCTATCCGTAGACCATGAAAGAGAGGGCTTGGTGGGG 171
```

QY 152 ACTTCTTGCAAGAAACAGATATGACATCAGACAGTACTCCGGCTTCGGTGAAGTTG 211
 DB 172 AGTTCCTAGAGACCCAGAAATATGATCTGCTTGGAAGTACCGCTTTGAGTACCTCAGCG 231
 QY 212 CTAGGTCGCCATTACCACTACCTGTTAGTCAATCTTTGGGAAGATCTACTCGGAA 271
 DB 232 TGACCTACGAGCCCATGAGCTATGATGCTGCTTACTTTGGTGAATGATCAGCACTCGGGA 291
 QY 272 CCGCGCCCTCAAGAGTTCAACGCTTCCTTGATGATGCTTCCTGATCTTCGGGTTCCCT 331
 DB 292 CTCACCCCAAGAACTTCTGCTGCTCTTTTGAACACCGGCTCTCCCAACTTGTGGGTCCT 351
 QY 332 CTATCTACTGCAAGAGCAATGCTGTCGAAGAACACCAAAAGATTCGATCCGAGAAAGTGT 391
 DB 352 CTGTCTATGCAAGAGCAGAGCTGTCACAGTCACTCCGCTTCAACCCAGCGAGTGT 411
 QY 392 CCACTTTCAGAACTTAGGCAAAACCTTGTCTATACACTAGGTTACAGGTTAGCATGCAAG 451
 DB 412 CCACCTTACTCAGCAACCAATGGGAGACCTTCTCCCTGAGATATGGCAGTGGCAGCTCACCG 471
 QY 452 GAATCTTAGGCTATGATACCGCTCACTGCTCCAAATGTTGACATTTGAACAGACAGTAG 511
 DB 472 GCTTCTTGGCTATGACACCTGACTGTCAGAGATCCAGGTCCCAACGAGAGTTG 531
 QY 512 GACTTAGACCCCAAGAACCAAGGTATGCTTCACTATGACAGAAATTCGATGGCATCTTG 571
 DB 532 GCTTAGAGAGAAATAGAGCTGTACCAACTTGTCTATGACGAGATTTGATGATGATCATG 591
 QY 572 GTATGGCATACCCATCGCTCGGTCAGAGTACTGATACCTGTTGTTGACAAATGATGA 631
 DB 592 GCCTGGCTTACCTGCTGCTGCTGATGAGGCCACCAACAGCTATGACAGGGCATGATGG 651
 QY 632 ACCGACACTATGATCTAAGAACTTGTCTGCTTATACATGACAGAAATGGCCAGAGGA 691
 DB 652 AGAAGGGGCGCTTCAACAGCCCGCTTCAAGCTTACTCAGCAACGAGGAGCTCCA 711
 QY 692 GCATGCTCAGCG-----TTGAGCTATTTGATTCATCTTACTACAGAGATCTTCTACT 745
 DB 712 GCGGGGAGCGGTTCTTTTGGGGGTGATGATGACGCTGTACAGGGGAGATCTTACT 771
 QY 746 GGGTTCACGATCACTGTGACAGAGTACTGAGCAATTCATCTGTGACAGTGTCACTCATCAG 805
 DB 772 GGGGCGCTGTACCCAGGAACCTTACTGAGATTTGGCATTTGAAGATTCCTCATCGGCG 831
 QY 806 GTGTGG---TTGTTGATGTTGAAGGTGATGCTAAGCTATCTTGAATACCGGTATGCTCCA 862
 DB 832 GCAAGGCTCGGGCTGTGTCTGAGGGTTGCGAGGGCATCGTGGACACAGGCACTCTGC 891
 QY 863 AGCTGCTGGAACCTTACAGAGCCGATTTCTCAACATTCAGCAAGCTATTTGAGACACAGGA 922
 DB 892 TGCTACTGTGCTCCAGCAGATGATGAGTCTTCTGAGGCGCACAGGGGCGCCAGGAGG 951
 QY 923 ACCAGTACGCTGATTTGATCATAGATTTGAGCAACCTTAGTACATGCTTACAGTTGCT 982
 DB 952 ATGATATAGACAGTTCTCTCGGAACCTGTAACAGATTCAGAAATCTGCCAGCTTGACT 1011
 QY 983 TTGAGATCAACGCGCAAGATGTAACCACTGACCCCTCGGCTTATACCAAGCCAGATCAAG 1042
 DB 1012 TCATCATCAATGTTGAGAGTTCCCTGCGCACCTTCTCTATATCTCAGTAAACAAG 1071
 QY 1043 GGTTCGACCAAGTATTCAGAGTGAAGAACATTC-----CAGAAAT 1087
 DB 1072 GCTACTGACCGTGGAGGTGAGCCCACTTACTGTCTCCAGAAACGCGCAGCCCTGT 1131
 QY 1088 GGAATCTTGGAGATGTTGATCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1147
 DB 1132 GGAATCTTGGAGATGTTGATCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1191
 QY 1148 TCGTTGGGCTAGCTTAAGCAATCT 1171
 DB 1192 GAGTAGGCTTTGCGACCTCGCGCT 1215

RESULT 12
 US-09-873-319-641
 ; Sequence 641, Application US/09873319A
 ; Publication No. US20030134324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Munge, William E.
 ; APPLICANT: Kulkarni, Prakash
 ; APPLICANT: Getzenberg, Robert H.
 ; APPLICANT: Muga, Iwao
 ; APPLICANT: Yamamoto, Jun
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
 ; FILE REFERENCE: 44921-5029-US
 ; CURRENT APPLICATION NUMBER: US/09/873.319A
 ; EARLIER FILING DATE: 2001-06-05
 ; EARLIER APPLICATION NUMBER: US 60/223,323
 ; NUMBER OF SEQ ID NOS: 755
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 641
 ; LENGTH: 1366
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20030134324A1 U75272
 US-09-873-319-641
 Query Match 21.2%; Score 248.4; DB 13; Length 1366;
 Best Local Similarity 53.8%; Pred. No. 1.1e-75;
 Matches 594; Conservative 0; Mismatches 486; Indels 24; Gaps 3;
 QY 92 TTCTCTCTACAAAGTATGATCTCTCGGTATAGGGGCTGGAAGAACTGATGATCTTCTAGAG 151
 DB 112 TGCCCCGTAAGAAATTTAAGTATCGGTGAGACATGAAAGAGGCTTGTGGGG 171
 QY 152 ACTTCTTGCAAGAAACAGATATGACATCAGACAGTACTCCGGCTTCGGTGAAGTTG 211
 DB 172 AGTTCCTAGAGACCCAGAAATATGATCTGCTTGGAAGTACCGCTTTGAGTACCTCAGCG 231
 QY 212 CTAGGTCGCCATTACCACTACCTGTTAGTCAATCTTTGGGAAGATCTACTCGGAA 271
 DB 232 TGACCTACGAGCCCATGAGCTATGATGCTGCTTACTTTGGTGAATGATCAGCACTCGGGA 291
 QY 272 CCGCGCCCTCAAGAGTTCAACGCTTCCTTGATGATGCTTCCTGATCTTCGGGTTCCCT 331
 DB 292 CTCACCCCAAGAACTTCTGCTGCTCTTTTGAACACCGGCTCTCCCAACTTGTGGGTCCT 351
 QY 332 CTATCTACTGCAAGAGCAATGCTGTCGAAGAACACCAAAAGATTCGATCCGAGAAAGTGT 391
 DB 352 CTGTCTATGCAAGAGCAGAGCTGTCACAGTCACTCCGCTTCAACCCAGCGAGTGT 411
 QY 392 CCACTTTCAGAACTTAGGCAAAACCTTGTCTATACACTAGGTTACAGGTTAGCATGCAAG 451
 DB 412 CCACCTTACTCAGCAACCAATGGGAGACCTTCTCCCTGAGATATGGCAGTGGCAGCTCACCG 471
 QY 452 GAATCTTAGGCTATGATACCGCTCACTGCTCCAAATGTTGACATTTGAACAGACAGTAG 511
 DB 472 GCTTCTTGGCTATGACACCTGACTGTCAGAGATCCAGGTCCCAACGAGAGTTG 531
 QY 512 GACTTAGACCCCAAGAACCAAGGTATGCTTCACTATGACAGAAATTCGATGGCATCTTG 571
 DB 532 GCTTAGAGAGAAATAGAGCTGTACCAACTTGTCTATGACGAGATTTGATGATGATCATG 591
 QY 572 GTATGGCATACCCATCGCTCGGTCAGAGTACTGATACCTGTTGTTGACAAATGATGA 631
 DB 592 GCCTGGCTTACCTGCTGCTGATGAGGCCACCAACAGCTATGACAGGGCATGATGG 651
 QY 632 ACCGACACTATGATCTAAGAACTTGTCTGCTTATACATGACAGAAATGGCCAGAGGA 691
 DB 652 AGAAGGGGCGCTTCAACAGCCCGCTTCAAGCTTACTCAGCAACGAGGAGCTCCA 711
 QY 692 GCATGCTCAGCG-----TTGAGCTATTTGATTCATCTTACTACAGAGATCTTCTACT 745

Db 712 GCGGGGAGCGGTTGCTTTGGGGGTGTGATAGCAAGCTGTACACGGGGCAATCTACT 771
Qy 746 GGGTTCCAGTCACTGTGACAGCACTACTGGCAATTCATGTGGA CAGTGTCAACATGCG 805
Db 772 GGGCGCTGTGACCCAGGAACCTACTGCGAGATTGGCAATGAAGTTCTCATGGCG 831
Qy 806 GTGTGG---TTGTTGATGTGAAGGTGATGTCAAGTATCTTGGATACCGGTACCTCA 862
Db 832 GCCAGGCTCCGGCTGGTGTTCGTAGGGTTGCCAGGCGATGTGACACAGGCACTCTC 891
Qy 863 AGCTGTGCGGACCTAGCAGGACATTCCTCAACATTCGCAAGGCTATTGGACCAAGCA 922
Db 892 TGTCTAATGTGCTCCAGCACTATGAGTGTCTCTCTGCAAGCCACAGGGCCCAAGAG 951
Qy 923 ACCAGTACGGTGAATTTGACATGATTTGCAACACCTTAGCTACATGCTACAGTTGTCT 982
Db 952 ATGAGTATGGAACAGTTCTCTGTGAACTGTAAACAGCATTCAGAAATCTCCAGCTTGA 1011
Qy 983 TTGAGATCAACGGCAAGATATACCCACTGACCCCTCCGCTTATACCAAGCATCAAG 1042
Db 1012 TCATCATCAATGTGTGGAGTTCCTCTGCAACCTTCCTCTATATCTCAGTAAACAG 1071
Qy 1043 GGTTCGCAACGATGATTCAGAGTGAACCAATTC-----CAGAAAT 1087
Db 1072 GCTACTGCAACCGTGGAGTGAAGCCCACTACCTGTCTCCAGAACGGCCAGCCCTGT 1131
Qy 1088 GGATCTTGGAGATGTGTTCATTCGTGAGTACTACAGCGCTTTGACAGGGCCAAACACC 1147
Db 1132 GGATCTCGGGGATGTCTCTCAGGTCCTACTATTCCTGCTACGATTTGGGCAACACA 1191
Qy 1148 TCGTTGGGCTAGCTAAAGCATCT 1171
Db 1192 GAGTAGGCTTTGGCACTGCCGCT 1215

RESULT 13
US-09-960-706-984
; Sequence 984, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Mungier, William E.
; TITLE OF INVENTION: Identifying drugs for and diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 984
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 U75272
US-09-960-706-984

Query Match 21.2%; Score 248.4; DB 13; Length 1366;
Best Local Similarity 53.8%; Pred. No. 1,1e-75;
Matches 594; Conservative 0; Mismatches 486; Indels 24; Gaps 3;
Qy 92 TTCTCTCTCAAGGTAATCTCTCCGTAAGGGCGTGAAGGAACATGACTTCTAGAG 151
Db 112 TGCCTCTGAAGAAATTTAACTATTCGTGAGACCATGAAGGAAGGGCTTCTGGGG 171
Qy 152 ACTTCTTGCAAGAACAGTATGAGCATGACAGCAAGTACTCCGGCTTGGTGAAGTTG 211
Db 172 AGTTCTGAGGACCCACAAAGTATGATCTCTTGGAAAGTACCGCTTTGGTACCTACGG 231

Qy 212 CTAGCGTGCACCTTACCACTACCTTGATAGTCAATCTTTGGGAAGATCTACTGGGA 271
Db 232 TGACCTACGAGCCCAATGAGCTTACATGATGCTGCCCTCACTTTGGTATGATACATGGGA 291
Qy 272 CCCCCTGCAAGATTCACCGTTCTCTTGATATGTTGTTCTCTGACTTCTGGGTTCCCT 331
Db 292 CTCACCCCAAGAACTTCCGTGTCTTTTGAACCGGCTCTCTCAACTTGTGGTCCCT 351
Qy 332 CTATCTTACGCAAGCAATGCTCTGCAAGAACCAACCAAGATTGATCCGAGAAAGTCT 391
Db 352 CTGTCTACTGCAAGAGCCAGGCTGTGACACAGTCACTCCGCTTCAACCCACAGAGTCT 411
Qy 392 CCACTTCCAGAACTTAGGCAAACTTGTCTATACATACATGAGTACAGTACCAAG 451
Db 412 CCACTTACTCCACCAATGGGCAAGACTTCTCTCTGATGATGGCAGCTTCAAC 471
Qy 452 GAATCTTAGGCTATGATACCTGTCACTGTCTCCAACTTGTGACATTCACAGACAGTAG 511
Db 472 GCTTCTTGGCTATGACACCTGACGTCTCAGAGCATTCAGGTCCCAACAGAGTTG 531
Qy 512 GACTTAGCAACCAAGAACAGGTATGTCTTCACTATGCAAGAAATTCATGCGATCTTG 571
Db 532 GCTTAGTGAATGAGCTGTGATCCAACTGTCTATGCGAGTTGATGGCATGATG 591
Qy 572 GATAGGCATACCCATGCTCGCTGAGAGTATCCGATACCTGTGTTGACAAATGATGA 631
Db 592 GCTGTGCTTACCTGTCTGTCTGTGATGAGGCCCAACACATGACAGGGCATGTGC 651
Qy 632 ACCGACACCTAGATGATCTCAAGACTTGTCTCGGTTTACATGACAGAGATGCGCAGAGA 691
Db 652 AGAAGGCGCCCTCACACAGCCCGCTTCAAGGCTTACTCAGCAACAGCAGGGCTCCA 711
Qy 692 GCATGCTCAAGC-----TTGAGCTATTGATTCATCTACTACTACACAGGATCTTCACT 745
Db 712 GCGGGGAGCGGTTGTTCTTTGGGGGTGTGATAGCAGCTGTACACAGGGGAGATCTACT 771
Qy 746 GGGTTCCAGTCACTGTGACAGCACTACTGGCAATTCATGTGACAGTGTCAATCATAGCG 805
Db 772 GGGCGCTGTGACCCAGGAACCTTACTGTGCAATGTGGCAATTAAGAGTTCTCATGAGCG 831
Qy 806 GTGTGG---TTGTTGATGTGAAGGTGATGTCAAGCTATCTTGGATACCGGATCGTCA 862
Db 832 GCCAGGCTTCGGCTGTGTGTTCTGAGGGTTGCCAGGCCATCTGTGACACAGGCACTCTC 891
Qy 863 AGCTGTGCGACCTAGCAGCAGCAATTCCTCAATTCAGCAAGCTATTGGACCAACAGA 922
Db 892 TGTCTAATGTGCCCCAGAGTACATGATGATGCTCTTCTGCAAGGCAACAGGGGCCAAGAG 951
Qy 923 ACCAGTACGGTGAATTTGACATGATTTGCGACAACTTAGCTTACATGCTTACAGTTGTCT 982
Db 952 ATGAGTATGGAACAGTTCTCTGTGAACCTGTAAACAGCAATTCAGAAATCTGCCAGCTGACCT 1011
Qy 983 TTGAGATCAACGGCAAGATGTAACCACTGACCCCTCCGCTTATACCAAGCATCAAG 1042
Db 1012 TCATCATCAATGTGTGAGTTCTCTCTGCACTTCTCTCTATATCTCAGTAAACAG 1071
Qy 1043 GGTTCGCAACAGTGAATTCAGAGTGAAGAACCATTC-----CAGAAAT 1087
Db 1072 GCTACTGCAACCGTGGAGTGAAGCCCACTACCTGTCTCTCCAGAACGGCCAGCCCTGT 1131
Qy 1088 GGATCTTGGAGATGTGTTCATTCGTGAGTACTACAGCGCTTTTGAAGGGCCCAACACC 1147
Db 1132 GGATCTCGGGGATGTCTCTCTCAGGTCCTACTATTCCTGTACAGACTTGGGCAACACA 1191
Qy 1148 TCGTTGGGCTAGCTAAAGCATCT 1171
Db 1192 GAGTAGGCTTTGGCACTGCCGCT 1215

RESULT 14
US-10-051-874-53
; Sequence 53, Application US/10051874
; Publication No. US2004000557A1


```

; APPLICANT: HAPALIA, April J.A.; LU, Dnyung Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Depoctraya; KALILICK, Deborah A.
; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKIDAR, Jayalaxmi
; APPLICANT: YANG, Vunming; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: BAUGHN, Mariah R.; BOROMSKY, Mark L.
; APPLICANT: SANJANMALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Nandinder K.
; APPLICANT: LAL, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 30
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 7477287CBI
US-10-274-639-30

```

```

Query Match      19.3%; Score 226.8; DB 12; Length 1173;
Best Local Similarity 54.4%; Pred. No. 3.9e-68;
Matches 605; Conservative 0; Mismatches 472; Indels 35; Gaps 6;

```

```

QY      83 TCACCGCATTCCTCTCAAGGTAAGTCTCTCCGTGAAGGCGTGAAGAAACATGAGAC 142
DB      50 TCTGCAGATCAATTGAGAGAAAGGCAAGCTATCCGCCAGAAATGAGAGACAGGGTG 109

QY      143 TTCTAGAAGACTTCTGCAAGAAAC-----AGATGCATCAGCAGAACTACTCCG 196
DB      110 TACTGAGAGCGTTTCTGAGGAGACACCAAGGCTGATCCAAATGGCAATATATTTC 169

QY      197 GCTTCGTAAGTGTCTAGAGGTGCACTTACCAACTACTTGATGATCAATACCTTTGGGA 256
DB      170 ATAATGATGCTGTGTCTTATGAGCCCTTACCAACTGATCTTTTACTTTGGGG 229

QY      257 AATTTACTCGGAACCCCGCTCAAGAGTTCAAGCTTCTTTGATGATGTTCTCTG 316
DB      230 AGATGCACTGGGACACCAACCCCAAAATTTC-CTAGTCTTTGATACGGGTTCTCCA 288

QY      317 A-CTTCTGGGTTCCCTCTATCTTACTGCAAGCAATGCTGCAAGAACCAAGATTC 375
DB      289 ATCTGTAGCTGCTCCCTCACTTACTGCAAGCAAGTGTGCTCCAAATCAACACAGTTC 348

QY      376 GATCCGAGAAAGTGTGCAACTTCCGAAGCTTGAAGCAACCTTGCTATATACACTAGG 435
DB      349 AATCCAGGCTGTCTCCACTTTCAGAAACGATGAGACAAACTTATGACTATCTTATGG 408

```

```

QY      436 ACAGTAGATGCAAGGAATCTTAAGCTATGATACCGGTCACTGTCTCAACATTGAGAC 495
DB      409 AGTGGAGCTGAGTGTGTCTTCTGGGCTATGACATGTAATCTGTTCATTAACATCTGTTC 468

QY      496 ATTCAACAGACAGTAGAAGCTTACCAACCAAGAACAGTAGATGCTTCACTATGAGAA 555
DB      469 AATTAACAGAGTGTGGCTGAGTAGAATGAGGCCAGGAGACCCCTTTACTATTTAGAC 528

QY      556 TTGATGATCTCTTGGATATGAGATACCATGCTGCGGTCAAGATATCTGATACCTGTG 615
DB      529 TTGAGGAGATCTGGGATGAGCTTACCAACATGAGAGAGGGAATTCCTTACAGTA 588

QY      616 TTGACAAATGATGTAACCGAATCTGATGATCTTGTCTCGTTTACATGAGAC 675
DB      589 ATGCAAGGAGATGTGACAGAGACCACTTACTCAGCCGTCTTACGCTTATCTTACCC 648

QY      676 AGGAATG-----CCAGAGAGATGCTCAAGCTTGAAGCTATTTGATTCATCTTAC 726
DB      649 TGCAGACCAACCGCCAGATTTGTGAGAGCTCATCTTGGAGTGTGACCCCAACTT 708

QY      727 TACACAGATCTTCTCACTGGGTTCCAGTCACTGTGACAGACTGAGCAATT--CACT 783
DB      709 TATTTGTCAGATCATCTGACCCCTGTCAAGCCGGAACGTGACTGACAGATTGCAATC 768

QY      784 GTGACAGTGTCAACATCAAGCGGTGTGTGTCATGTGAAGTGTGATCAAGCTATC 843
DB      769 GAGGAATTTGCCATGCTGTAAACAGGCACTGTGTGTCTGAGGGTTGCCAGGCAATT 828

QY      844 TTGATACCGGTAACGTCAAGCTGTGCGGACCTTACAGAGCACTTCAACATTACGAA 903
DB      829 GTGATACCGAAGACCTTCTGCTGAGATTTCTCAGACAGTAATGCTCTTCTGCAAG 888

QY      904 GCTATTGAGGCCACACAGAACAGTAAGGTGATTGACATGATTCGACAAACCTTAGC 963
DB      889 GCAACAGGACCCGAGAGGCTCAAGATGTGACTTTGTGATCAATGACAGTACATACAG 948

QY      964 TACATGCTCAAGTGTCTTTGATGATCAACGCGAAGATGTACCATCCATGACCCCTCGCC 1023
DB      949 AGCATGCCCAATCATCTTCAATCATCGCGGGGCCAGATTTCTTGTGCTCTCTTGAA 1008

QY      1024 TATACAGCCAGGATCAAGGTTCTGACAGTGTTCAGAGTTCAGAGTGAACCATTC--- 1080
DB      1009 TATGTTTCAATTAACATGCTTACTGACAGCTTGAAGTGAAGCCACTGCTGCTCTCC 1068

QY      1081 -----CAGAAATGATCTTGGAGATGTGTTCATTCGTGATGATCAACGCTC 1128
DB      1069 CGCAGTGGGAGGCGCCCTCTGATTTCTGGGGATGTCTTCTCAAGGAATATGCTGTGC 1128

QY      1129 TTTGACAGGCGCAACCACTCGTTGGGCTAGC 1160
DB      1129 TATGATGCGCAACCAACAGGGTGGGCTTTGCG 1160

```

Search completed: February 5, 2004, 06:00:52
Job time : 2429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 01:39:30 ; Search time 2886 Seconds
(without alignments)
9878.431 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173

Sequence: 1 atgaactccttaagctctt.....ggctagctaaagcaatctga 1173

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	401.6	34.2	933 13	BQ219820 AGENCOURT
2	387.2	33.0	472 12	BG938320 IABO15E12
3	358	30.5	631 9	A1892188 mJ84C05.Y
4	354.8	30.2	430 12	BG937697 IABO05E01

5	337.2	28.7	727 9	A1326975
6	323.8	27.6	399 12	BG937723
7	312.6	26.6	410 14	CB220875
8	300.8	25.6	1347 11	AK004109
9	300.8	25.6	1408 11	AK053965
10	300.6	25.6	383 12	BG938086
11	288.6	24.6	623 13	BQ417551
12	271.8	23.2	910 14	CD558204
13	268.8	22.9	826 14	CD516662
14	266.2	22.7	1101 14	CD499194
15	266	22.7	1083 14	CD503881
16	264.4	22.5	1110 14	CD510280
17	264.2	22.5	678 14	BM783853
18	262.6	22.4	1124 14	CD504528
19	262.2	22.4	1157 14	CD496827
20	262	22.3	479 9	AA028632
21	262	22.3	665 12	BM795763
22	262	22.3	691 12	BM817889
23	261.6	22.3	1155 14	CD496442
24	257.6	22.0	605 13	BQ477584
25	257	21.9	1169 14	CD510281
26	256.8	21.9	1122 14	CD509091
27	254.8	21.7	655 12	BM782932
28	254.8	21.7	1224 14	CD505876
29	254.4	21.7	1139 14	CD508825
30	253.4	21.6	443 14	CB220074
31	252	21.5	651 12	BM783855
32	251.4	21.4	1199 14	CD509274
33	251.2	21.4	1197 14	CD505579
34	251	21.4	1140 14	CD497655
35	250.6	21.4	568 12	BQ027751
36	249.6	21.3	505 9	A1479358
37	249	21.2	1134 14	CD498219
38	248.2	21.2	1110 14	CD508053
39	247.8	21.1	307 14	CB221243
40	247.2	21.1	1154 14	CD503882
41	246.4	21.0	297 14	CB220520
42	246	21.0	546 9	A1324867
43	245.2	20.9	1175 14	CD508748
44	243.6	20.8	450 4	BX525462
45	242.4	20.7	517 9	A1385490

ALIGNMENTS

RESULT 1
LOCUS BQ219820 933 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT 7578073 NCI CGAP_Sci Mus musculus cDNA clone
IMAGE:6051528 5', mRNA sequence.
ACCESSION BQ219820
VERSION BQ219820.1 GI:20401220
KEYWORDS EST
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS BUKARYOLCA, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LLM13305 row: 1 column: 01
High quality sequence stop: 691.

Query Match	30.5%	Score 358;	DB 9;	Length 631;
Best Local Similarity	73.0%;	Pred. No. 1.1e-92;		
Matches 460;	Conservative	0;	Mismatches 170;	Indels 0;
Gaps	0;			
Db	301	AACATCCAGCAGGCCATGTGAGCCACACAGAACCAAGTCAGTGAATTTGATGCATGCATGC	360	
Qy	952	GACAACTTACTTACATGCTCCTTACAGTGTCTTTTGATGATCAACGGGAAGATGTACCATCTG	1011	
Db	361	GACAACTTACTTACATGCTCCTTACAGTGTCTTTTGATGATCAACGGGAAGATGTACCATCTG	420	
Qy	1012	ACCCCTCGGCTATACACGAGGATCAAGGGTTCTGCACAGTGATGCTC	1063	
Db	421	ACCCCTCGGCTATACACGAGGATCAAGGGTTCTGCACAGTGATGCTC	472	
RESULT 3	Al1892188			
LOCUS	Al1892188			
DEFINITION	Al1892188	631 bp	mRNA	linear
ACCESSION	U094G05.Y1	Scars mouse p33MF19.5	Mus musculus	CDNA clone
VERSION	IMAGE:482792.5	similar to TR:Q28950	Q28950	PREPROCHYMOSIN
KEYWORDS	PRECURSOR ;	RNA sequence.		
ORGANISM	Al1892188			
SOURCE	Al1892188.1	GI:5598090		
REFERENCE	EST.			
AUTHORS	Mus musculus	(house mouse)		
TITLE	Unpublished			
JOURNAL	Contact: Marra M/WashU-NCI Mouse EST Project 1999			
COMMENT	Washington University School of Medicine			
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: mouseest@watson.wustl.edu			
	This clone is available royalty-free through LML ; contact the			
	IMRG Consortium (info@image.lml.gov) for further information.			
	This read is a RESEQUENCE of a previously sequenced mouse clone			
	This read has been verified (found to hit its original self in the			
	correct orientation)			
	Possible reversed clone: similarity on wrong strand			
	Seq primer: -40RP from Gibco			
	MCI:293536			
FEATURES	High quality sequence stop: 479.			
SOURCE	Location/Qualifiers			
	1. .631			
	/organism="Mus musculus"			
	/mol type="mRNA"			
	/db xref="taxon:10090"			
	/clone="IMAGE:482792"			
	/dev_stage="19.5 dpc total fetus"			
	/lab_host="DH10B (ampicillin resistant)"			
	/clone_lib="Scars mouse p33MF19.5"			
	/note="Vector: pRTT3 (Pharmacia) with a modified			
	polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA			
	was primed with a Not I - oligo(Td) primer [5'			
	TGTTTCAATCATCTGAAGTGGAGCGGCCGCGCATTTTCTTTTCTTTTCTTTT 3']			
	double-stranded cDNA was size selected, ligated to Eco RI			
	adapters (Pharmacia), digested with Not I and cloned into			
	the Not I and Eco RI sites of a modified pRTT3 vector			
	(Pharmacia). Library went through one round of			
	normalization to a Cot = 5. Library constructed by Bento			
	Soares and M.Fatima Bonaldo. RNA was kindly provided by			
	Dr. Minoru Ko (Wayne State University)."			
BASE COUNT	147 a	174 c	155 g	155 t
ORIGIN				

QY	228	CAACTACCTTGATGATGTGAATTAATCTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGAGTT	287
Db	1	CAACTACTCTGATATGAGAGATCTTTTGGACGATCTTACTTGGACACCAACCGCAGAGTT	60
QY	288	CACCGTTCCTTTGGATATCTGGTCTCTGACTTCTGGGATTCCTCTATCTACCTGCAAG	347
Db	61	CACCGTGTGTTTTTATACGGCTCTCTGAAACTCTGGGATACCTCTGTCTATCTGCAACAG	120
QY	348	CAATGCGCTGCAAGAACCAACAAAGATTGCATCCGAGAAAGTCGTCACCTTCCGAAGCTT	407
Db	121	CAAGGTGTGCGGAACCAACACCGTTTTGACCCATCCAAAGTCATCACTTTCGAAACT	180
QY	408	AGGCAAAACCCCTGTGCTATACACTACGCGGTACAGTAGCATGCAGAAATCTTAGGCTATGA	467
Db	181	GAGGAAGACCCCTGTTTGTCCAGTATGATCTGTGTAATGGAAGGCTTCTCGGCTTACGA	240
QY	468	TACCGTCACTGTCTCCAAATTTGTGACATTTGACACAGACAGTAGAGACTTTAGACCCAGA	527
Db	241	CACGTGCACAGTCTCTGATATATGATGTGCCCATCAGACTGTGGGCTGATGCCAGGA	300
QY	528	ACCAAGGATGATCTTTCACCTATGAGAAATTCGATGCGCATCTTGGTATGAGATACCATC	587
Db	301	ACCGGGGGAATCTTCACTCACTCTCCATTTGAAGGATCTCGGCGCTGATCTTCAAC	360
QY	588	GCTGCGCTCAGAGTACCTGATACCTGTGTTTGAACAATGATGAACCAACCACTAGTAGC	647
Db	361	TTTGTGCTCCAAATCTCAGTACCCATATTTGACAAATGATGAACAGGACCTGGTGGC	420
QY	648	TCAAGACTTGTCTTCGGTTTATCATGACAGGAATGGCCAGGAGAGCATGCTCACGCTTGG	707
Db	421	CCAAGACTGTGTCTCCGTTTACATGACAGGAATGAGCAAGGCGAGCATGCTCACACTGGG	480
QY	708	AGCTATTGATTCATCTCTACTACACAGATCTCTTCACTGGGTTTCAGTCACTGTGCAGA	767
Db	481	GCGCATGATCAGTCCCTCACTTCAAGGCTCACTGCACCTGGGTGGCTGTGACAGTACAGGG	540
QY	768	GTACTGGCAATTCACCTGTGACAGTGTACACATCAGCGGATGTGGTTGTGCATGTGAAG	827
Db	541	ATATTGGCAGTTCCAGATGACAGAGATCAATTAATGTGAAGTGTGCTTGTCCAGG	600
QY	828	TGATGTCAAGCTATCTTGGATACCGGTAC	857
Db	601	TGGCTGCCCTGCCGTTCTGGACACAGGCAC	630
RESULT 4			
LOCUS	BG937697	430 bp	mRNA linear EST 11-JUN-2001
DEFINITION	1b005E01 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA		
ACCESSION	BG937697		
VERSION	BG937697.1	GI:14337069	
KEYWORDS	EST.		
SOURCE	Bos taurus		
ORGANISM	Bos taurus (cow)		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.		
AUTHORS	1 (bases 1 to 430)		
TITLE	Moore,S.s., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.		
JOURNAL	cDNA's from bovine abomasum tissue		
COMMENT	Unpublished		
	Contact: Dr. Stephen Moore		
	Beef Genomics Laboratory		
	Dept of AFNS, University of Alberta		
	410 Agr1/Fox, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada		
	Tel: 780 492 0169		
	Fax: 780 492 4265		
	Email: stephen.moore@ualberta.ca		
	The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)		
	trna1 in main database at high score of 844.0 and E-value of 0.0		
	PCR Primers		

FORWARD: M13 Forward
 BACKWARD: M13 Reverse
 Seg primer: T3 primer
 High quality sequence stop: 430
 POLYA=NO.

FEATURES	Location/Qualifiers
source	1. .430

BASE COUNT ORIGIN	97 a	130 c	119 g	84 t
----------------------	------	-------	-------	------

Query Match	30.2%;	Score 354.8;	DB 12;	length 430;
Best Local Similarity	89.1%;	Pred. No. 7.4e-92;		
Matches 383; Conservative	0;	Mismatches 47;	Indels 0;	Gaps 0;

QY	605	CGATACCTGTGTTTGACCAACATGATGAACGACACCTAGTAGTCAAGACTGTCTCGG	664
Db	1	CGATACCCGCTTTTGACCAACATGATGAACGACACCTGTGCCCCAAGACTGTCTCGG	60
QY	665	TTTATACATGACAGGAATGGCCAGAGAGACATGCTCAGCCTTGGAGCTATTTGATCATCTT	724
Db	61	TTTATACATGACAGGAATGGCCAGAGAGACATGCTCACCCTGGGGGCCATTCGACCCGTCCT	120
QY	725	ACTACACAGGATCTCTTCACTGGGGTTCAGTGCACGTGCACAGACAGTACAGGCAATTCACTG	784
Db	121	ACTACACAGGATCTCTTCACTGGGGTTCAGTGCACGTGCACAGACAGTACAGGCAATTCACTG	180
QY	785	TGGACAGTGTCAACCATCAGCCGCTGTGTTGTTCATGTGAAGAGTGGATGTCAAGCTATCT	844
Db	181	TGGACAGTGTCAACCATCAGCCGCTGTGTTGTTCATGTGAAGAGTGGATGTCAAGCTATCT	240
QY	845	TGATATACCCGTAAGTCCAAAGCTGGTCGGAACCTAGCAGCGACATTTCTCAACATTACGCAAG	904
Db	241	TGACACAGGGGCACTCCCAAGCTGTGTGGGGCCACGACGAGGACATCTCTCAACATCCGACAGG	300
QY	905	CTATTGGAGCCACACAGAACCGATACGGTGTGATTTGACATGATTTGGACACAACTTAGCT	964
Db	301	CCATTGGAGCCACACAGAACCGATACGGTGTGATTTGACATGATTTGGACACAACTTAGCT	360
QY	965	ACATGACCTACAGTTGTCTTTTGAGATCAACGGCAAGATGTACCCACTGACCCCTCGGCT	1024
Db	361	ACATGACCTACAGTTGTCTTTTGAGATCAATGGCAAAATGTATCCCACTGACCCCTCGGCT	420
QY	1025	ATACCAAGCCA 1034	
Db	421	ATACCAAGCCA 430	

RESULT 5	LOCUS	DEFINITION
AI326975/c	AI326975	727 bp mRNA linear EST 23-DEC-1998
	nm181e03.x1	Soares mouse p3jM9F19.5 Mus musculus cDNA clone
	IMAGE:482524 3'	similar to TR:Q28075 Q28075 CHYMOSIN C. ;, mRNA

ACCESSION	AI326975
VERSION	AI326975.1
	GI:4061404

SOURCE ORGANISM	REFERENCE AUTHORS
Mus musculus (house mouse)	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 727)	
Werra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.	

TITLE The WashU-HIMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
Geisels, S., Kucaba, T., Lacey, M., Le, M., Matlin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theisling, B., Wylie, T., Lennon, G., Soares, B., Watson, R. and
Waterston, R.

FEATURES	Location/Qualifiers
source	1. .727

BASE COUNT	178 a	180 c	195 g	174 t
ORIGIN				
Query Match	28.7%	Score 337.2;	DB 9;	Length 727;
Best Local Similarity	71.4%	Pred. No. 1,38-86;		
Matches 474;	Conservative	0;	Mismatches 183;	Indels 7;
				Gaps 2

Qy	511	GGACTTGGACCCCAAGAACCGGGATGATCTTCACTTACCTATGACGAATTGATGGATCCTT	570
Db	727	GGCTTGATGACCCAGAAACGGGGGACATCTTAACTTACCTTCCATTTGATGGCATCTTG	668
Qy	571	GGATVGGCATACCATGCTCGCGTCAGAGTACTCGATACCTGTGTGTGAACAATGATG	630
Db	667	GGCTTGCCAACTCTTCTTTTGCTCCAAATACTCAGTACCATATTTGACAACTGATG	608
Qy	631	AACCGAACCTAGTAGTCAAGATTGTTCCTCGSTTTACATGACAGGAATGGCCAGAG	690
Db	607	AACAGGCACTGGTGGCCCAAGACTGTCTTCGSTTTACATGACAGGAATGACGAGGG	548
Qy	691	AGCATGCTCAAGCTTGGAGCTATTGATCCATCCTACTACACAGGATCTCTTCATCGGTT	750
Db	547	AGCATGTCTACACTGGGGGCCATGATCAGTCCACTTCAATAGGTCACTGCATCGGGTG	488
Qy	751	CCAGTCACTGTGCAGCACTACTGGCAATTCACTGTGACAGTGTCAACCATCAGCGGTG	810
Db	487	CCGTGTGACAGTACAGGGATATTGGCAGTTCACAGTGCACAGATTCACATCAATGTGAA	428
Qy	811	GTTGTTGCATGTGAAGGTGATGTCAAGCTACTTGGATACCGGATGCTCAAGGTGTG	870
Db	427	GTGTGTGCTGTGCAAGGTGGCTGCTGCCGTCTTGACACAGGCACTGCCCTGTTGACG	368
Qy	871	GGACCTT-AGCAGCGACATTTTCAACATTTACAGCAAGTATTGGAGCCACACGAAACAGTA	929

Db 367 GGGCTGTATTAAGACATCTCAATATTCAGCAAGTCATTTGAGCTGTCAGGACCATTA 308
QY 930 CGGTGATTTGACATAGATTTGCGACACCTTAGTACATGCTTACAGATTTGTTGAGAT 989
Db 307 TGACCAATTTGACATGCTGCTGAGAGCTTGACATCATGCCCCACCGTTTCTTTGAGAT 248
QY 990 CAACGGCAAGATGTACCACTGACCCCTCCGCTATACACAGCAGATCAAGGGCTTCG 1049
Db 247 CCATGATGAGGAGTTCCCACTGCAACCCCTAAGCTTATACCAACAGGTCAGGGCTTCG 188
QY 1050 CACCACTGATTTCCAGAGTGAACACCATTTCCAGAAATGATCTTGGAGATGTTTCAT 1109
Db 187 CTCAGTGGCTTC-----AAGCAGGGCTCCCACTGTGATCTCTGGGAGATGTTTCAT 134
QY 1110 TCGTAGTACTACAGGCTTTTGACAGGGCCAAACACCTGTTGGGCTAGCTTAAGCAT 1169
Db 133 TCGGAGTTTCTACAGTGTGTTTGACAGACCAACATGTTGGGCTGCGCAAGGCTAT 74
QY 1170 CTGA 1173
Db 73 CTGA 70

RESULT 6
BG937723 399 bp mRNA linear EST 11-JUN-2001
LOCUS 1AB005D06 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
DEFINITION
ACCESSION BG937723
VERSION BG937723.1 GI:14337095
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 399)
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
CDNA's from bovine abomasum tissue
Unpublished
JOURNAL
COMMENT
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
The sequence best matches gb:BFU19786 (Bos primigenius prothymosin
mRNA, complete cds) in main database at high score of 767.0 and
E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 399
POLYA-No.

FEATURES
source Location/Qualifiers
1..399
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMRF"-strain
/clone_lib="Bovine Abomasum cDNA Library"
/note="Organ: Abomasum; Vector: Uni-Z2APK; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT 89 a 121 c 112 g 77 t
ORIGIN
Query Match 27.6%; Score 323.8; DB 12; Length 399;

Best Local Similarity 88.2%; Pred. No. 7.3e-83;
Matches 352; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 662 CGGTATCATGACACAGAAATGGCCAGAGAGATCTCAACCTTGGAGCTATTGATCAT 721
Db 1 CGGTATCATGACACAGAAATGGCCAGAGAGATCTCAACCTTGGAGCTATTGATCAT 60
QY 722 CCTACTACAGAGATCTCTTCACTGGGTCCAGTCACTGTGAGAGATCTAGGCAATTCA 781
Db 61 CCTACTACAGAGATCTCTTCACTGGGTCCAGTCACTGTGAGAGATCTAGGCAATTCA 120
QY 782 CTGTGACAGTGTACCATCAACGAGGCTGTGTTGTCATGTGAGAGTGTGATCAAGCTA 841
Db 121 CTGTGACAGTGTACCATCAACGAGGCTGTGTTGTCATGTGAGAGTGTGATCAAGCTA 180
QY 842 TCTTGATATCCGGTACCTGTCAGAGCTGTGTGACCTAGACAGCACTTCTCAATTCAC 901
Db 181 TCTTGATATCCGGTACCTGTCAGAGCTGTGTGACCTAGACAGCACTTCTCAATTCAC 240
QY 902 AAGCTATTGAGGACCAACAGACAGTACGGTGTGATGACATGATTTGCGACACCTTA 961
Db 241 AAGCTATTGAGGACCAACAGACAGTACGGTGTGATGACATGATTTGCGACACCTTA 300
QY 962 GCTACATGCTTACAGTGTCTTTGAGATCAACGGCAAGATGTACCCACTGACCCCTCCG 1021
Db 301 GCTACATGCTTACAGTGTCTTTGAGATCAACGGCAAGATGTACCCACTGACCCCTCCG 360
QY 1022 CCTATACAGCCAGGATCAAGGGTTCTGCACCAGTGCAT 1060
Db 361 CCTATACAGCCAGGATCAAGGGTTCTGCACCAGTGCAT 399

RESULT 7
CB220875 410 bp mRNA linear EST 10-FEB-2003
LOCUS 1AB027D10 Bos taurus Abomasum #1 library Bos taurus cDNA, mRNA
DEFINITION
ACCESSION CB220875
VERSION CB220875.1 GI:28291389
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 410)
Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
P.M.K. and Moore,S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished
JOURNAL
COMMENT
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
Insert Length: 410 Std Error: 0.00
POLYA-No.

FEATURES
source Location/Qualifiers
1..410
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Smooth muscle"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMRF"-strain
/clone_lib="Bos taurus Abomasum #1 library"
/note="Organ: Abomasum; Vector: Uni-Z2APK; Site_1: EcoR
I; Site_2: Xho I"
BASE COUNT 90 a 123 c 120 g 77 t
ORIGIN

/db_xref="taxon:10090"
 /clone="1110035E17"
 /issue_type="whole body"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="18-day embryo"
 40..1203
 /note="unnamed protein product; pepsinogen F
 (MGJ|MGJ:1915935, GB|NM_021453, evidence: BLASTN, 9%,
 match=1210)
 putative"
 /codon_start=1
 /protein_id="BAB23174.1"
 /db_xref="GI:12835164"
 /db_xref="MGJ:1915935"
 /translation="MKMLVGLVALSGLVIRKIPMKIKSMRENRESOVLDKYLEKY
 PRSAHVLEQRNRYATYEPERNYLDVYIGISIGPPRPFRLVDLTSGLVWPS
 IYSSPFAHHKAFNPSTFLVSGRPVNAVSGSGFLADYVLDITVVAQA
 FGLSLKEPFIEMEVAVFDGILGLGVNGLQGITPEVFNLMQLIPNLPAFYLSK
 DEKSGMLMGVDPSTYHGLHWVPSKPSYQLADSIKMGVEVIAADGCGQIMDT
 GTSLLTGRSSIVNIQNLIGAKASGDGEYFLKCDITNLPDVIPTIGSVTPVPSAY
 IRKRSNCRSNFSEGMDDPSDPEMWLGDVFLRYTFVPRANNRIGLAIPA"
 1332..1337
 /note="Putative"
 polyA_site
 1347
 /note="putative"
 BASE COUNT 296 a 364 c 358 g 329 t
 ORIGIN

Query Match 25.6%; Score 300.8; DB 11; Length 1347;
 Best Local Similarity 58.3%; Pred. No. 7.4e-76;
 Matches 574; Conservative 0; Mismatches 392; Indels 18; Gaps 2;

208 GTTGCTACGCGCCACTTACCACTACTGATGATCTTGGGAAGTCTACCTC 267
 220 GTAACCTATGAGCCCTATGAGAACTCAAGCTGCTTACATGGATCAAGCAAT 279
 268 GGAACCCGCGCTCAAGAGTTCACCGTCTCTTGATGATGTTCTGACTTCTGGGT 327
 280 GGACGCGCCCTCAGAGATTCAGGTTGCTTGATCTGATCTTCACTCTGAGGTA 339
 328 CCGCTATCTACTGCAAGCAATCCCTGCAAGAAACCAAGATTCGATCGAAGAA 387
 340 CCATCATCTATGCTCCAGCCGCTGCGCTCAACCAAGCCCTTCAACCGGTTGG 399
 388 TCGTCCACCTTCAGAACTTAGGCAACCTGCTTATACATAGGTAAGTACATG 447
 400 TCTTCACCTTCTCGTCTCAGGCGACCTGTGAATGTTGCTTATGCTCAGAGATG 459
 448 CAAGGATCTTACGCTATGATACGCTGCTCCACATTTGAGACATTCAGAGCA 507
 460 TCCGATTTCTTGCTATGACACTGTACGATTTGGGACCTCAGCGTTGGCCAGGC 519
 508 GTAGGACTTAGGACCCAGAAACAGAGTGTCTTACCTATGAGAAATTCAGTGCATC 567
 520 TTGGCGTGAAGCTGAGAACTTCTGCAATTTTCAAGATATGCTGTGATGATG 579
 568 CTGGTATGCTATCCCATCGTCCGCTGACAGTACTGATACCTGTGTTGACCAATG 627
 580 CTGGGCGTGGATACCCCAACCTTGCCCTTCAAGGAATCACCCGCTTTTGACCACTG 639
 628 ATGACCGACACCTTAGTACTCAAGCTTGTCTCGGTTTACATGAGACG---GAATGC 684
 640 TGGCTACAGGCTCTATCCCGAGAAATCTCTTGCTTACTTACGACGAAAGATGAA 699
 685 CAGAGACATGCTCAGCTTGAAGCTATGATTCATCTTACAGACAGATCTCTTAC 744
 700 AAGGCGAGATGCTAGTGTAGGAGTGAATCCCTCTTACACAGAGAGATTAC 759
 745 TGGGTTCCAGTACTGTGACAGATGCTGCAATTCCTGTGAGAGTGCATCAGTAC 804
 760 TGGGTACAGGTGTCCAAAGCCAGCTACTGCGCAATTAGCTGTGATGAGCATCTCCATGAT 819
 805 GGTGTGTTGTTGATGTGAAGGTGATGTCAAGCTATCTTGATACCGGTACGTCAG 864

Db 820 GGGGAGGCTATGCTGCTGATGATGCGCTGCAAGATATTATGACACAGGACCTCTTG 879
 Qy 865 CTGCTGGACCTTAGCAGGACATTTCTCAACTTAGCAGAGCTATTGGAGCCACAGAAC 924
 Db 880 CTGACCGGCGCCCGAAGCTCATGCTTAACTCAAGAACTTAATTTGGTCAAGGCTTCT 939
 Qy 925 CAGTACGGTGAATTTGACATGATTTGACACCTTACTACATGCTTACAGTTGCTTT 984
 Db 940 GGTACGCGAGTACTTCTTCAAGTGTACACCATCAACCCCTGCTGATTTGCTTTC 999
 Qy 985 GAGATCAACGCGCAAGATGTACCCACTGACCCCTCCGCTTATACAGCCAGATCAAGGG 1044
 Db 1000 ACCATGCGCAGGTACTACTTACCCAGTGCAGAGTGCCTTACATCCGAAAGATGCTCA 1059
 Qy 1045 TTTCGACCACTGATTTCAAGAGTGAAC-----CATTTCCAGAAATGG 1089
 Db 1060 CACAAATTCAGAGAGCACTTTGAGAGAGGCAATGATGACCATGACACCTGAGATGTTG 1119
 Qy 1090 ATCTTGGGAGATGTGTTTCAATTCGTGAGTACTACAGCGCTTTGACAGGCGCAACACTC 1149
 Db 1120 GTGCTGGGAGATGCTTCTTCTGAGCGCTGATTTACCGGTGTTGATGCGGCAATATACAG 1179
 Qy 1150 GTTGGGCTAGCTTAAAGCAATCTGA 1173
 Db 1180 ATTGCTGTGCTCTCTGCTGATGA 1203

RESULT 9
 AK053965
 LOCUS
 DEFINITION
 Mus musculus 2 days pregnant adult female oviduct cDNA, RIKEN
 full-length enriched library, clone: E23008A19 product: pepsinogen
 F, full insert sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 AK053965
 AK053965.1 GI:26343894
 HTC; CAP trapper.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2 99279253
 103496636

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 3 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komu, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Komu, H., Adachi, D., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komu, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M.,
 Suni, N., Ishii, Y., Nakamura, S., Hazama, T., Nishino, T., Harada, A.,
 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

REFERENCE
 AUTHORS
 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Komu, H., Adachi, D., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,

Kadota, K., Matsumura, H., Aeburner, M., Batalov, S., Casavant, T., Fleschmann, M., Gaasterland, T., Gissi, C., King, B., Koonin, H., Kuenli, P., Lewis, S., Matsuo, Y., Nakai, I., Pesole, G., Quackenbush, J., Schiraldi, L.M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furumori, M., Aono, H., Baldarelli, P., Barash, G., Blake, J., Boffelli, D., Bojunga, N., Carinci, R., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauer, P., Nordone, P., Ring, B., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilting, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohseki, S., and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection

JOURNAL
Nature 409 (6821), 685-690 (2001)

MEDLINE
21085660

PubMed
11217851

REFERENCE
5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
6 (bases 1 to 1408)

ADACHI, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furumori, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaku, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE
Direct Submission

JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers

source
1. 1408
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:E330008A19"
/db_xref="taxon:10090"
/clone="E330008A19"
/sex="female"
/tissue_type="oviduct"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
102..1265
/note="unlabeled protein product; pepsinogen F (MGI:1915935, GB|NM_021453, evidence: BLASTN, 99%, match=1210)
putative"
/codon_start=1
/protein_id="BAC35604.1"

/db_xref="GI:26343895"
/translation="MKMLVGLVALSEGLVPIPLMKIKSRENRESOVLYKDYLEK
PSRAHVLLEGRNPAVTEPERNVLDTYGIISIGPPQFFRVLDGSSVLWPS
ITCSSPACHNHAFPLRSSTPLVSGRPNNVYSGSEMSGLAYDVRGDLTVVAQA
FELSEEPFIEMVEYFDGILGLGPNLQGITTFVDFNLQGLIPQILFAYFYSK
DEKSLMGLGVDPSYGEHLHFWPSKSYQLVDSISMKGEVAVACGGCGIMDT
GSLITGPSSTLVNLIQNLIGASGDGEYFLKCDITNTLPIDVITIGSATVPSAY
IKRDSHNRKSRFPERGMDPSPEMVLGVLRILRYFVDPANRIGLAPA"

polya_signal
1393..1398
/note="putative"

polya_site
1408
/note="putative"

BASE COUNT
314 a 380 c 373 g 341 t

ORIGIN

Query Match 25.6%; Score 300.8; DB 11; Length 1408;
Best Local Similarity 58.3%; Pred. No. 7.6e-76;
Matches 574; Conservative 0; Mismatches 392; Indels 18; Gaps 2;

QY 208 GTTGCTAGCGTGCCTTACCACTACCTTGTATGATCACTTGGGAAGATCTACCTC 267
DB |||||
DB GTAACTTATGAGCTTATGAGAACTTACCTAGCTGTGTATGATCACTACCTAT 341
QY 268 GGAACCCCGCTCAAGATTTCAACCGTTCTTGTATGATGTTCTTCACTTGGGTT 327
DB |||||
DB GGACGCGCCCTCAGAGTTCAAGGTTTCTTGTATGATGATCTTCACTTGGGTT 401
QY 328 CCCTTATCTACTGCAAGATTCATGCTTCAAGAACCAACCAAGATTCGAGAAAG 387
DB |||||
DB CCACTCATATTTGCTCCAGCCCAAGCTCGCTGCTACCAACCAAGCTTCAACCGCTTGG 461
QY 388 TCGTCACTTCCAGAACTTGAAGCAACCTTGTCTATACCTACCTACCTGATGATG 447
DB |||||
DB TCTTCCACTTCTGCTTCCAGCCCACTGGAATGTTGCTTATGCTTCAAGAGATG 521
QY 448 CAAGAACTTGGGCTATGATACCGTCACTGTCTCAACATTTGATGATTCACAGACA 507
DB |||||
DB TCCGATTTCTTGTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 581
QY 508 GTAGACTTATGACCAACCAAGACAGGTGATGCTTCACTATGATGATGATGATGAT 567
DB |||||
DB TTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 641
QY 568 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 627
DB |||||
DB CTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 701
QY 628 ATGACCGACCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684
DB |||||
DB TGGCTTCAAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 761
QY 745 TGGCTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 804
DB |||||
DB TGGCTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 881
QY 805 GGT 864
DB |||||
DB GGGGAGGTATGCTGT 941
QY 865 CTGT 924
DB |||||
DB CTATACCGGCGCGGAGGCTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1001
QY 925 CAGTACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 984
DB |||||
DB GGTGAGGCGGAGTATCTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1061
QY 985 GAGTACGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1044
DB |||||

Db 1062 ACCATCGGACGTGTTACCTACCCAGTCCAGCCAGTGTCTACTACCCGAAGATCGTCA 1121
 QY 1045 TTCTGCACACAGTGAATTCAGATGAGAAC-----CATTCGCAAGAAATGG 1089
 Db 1122 CACAAATTCAGAGAGCACTTTGAGAGGGCAGTATGACCAATCAGACCTGAGATGTGG 1181
 QY 1090 ATCTTGAGAGATGTTGTTTATTCGTGAGTACTACACCGCTCTTTGACAGGGCCAAACCTC 1149
 Db 1182 GTGCTGGGGAGATGTTCTCTGAGGCTGATTTTCAACGTTGATCGGGCAATTAACAG 1241
 QY 1150 GTTGGGCTAGCTAAAGCAATCTGA 1173
 Db 1242 ATTGCTGCTCTCTCTGTCATGA 1265

RESULT 10
 EG938086 383 bp mRNA linear EST 11-JUN-2001
 LOCUS 1A011A08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
 DEFINITION
 ACCESSION EG938086
 VERSION EG938086.1 GI:14337458
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 383)
 Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
 CDNA's from bovine abomasum tissue
 JOURNAL Unpublished
 COMMENT Contact: Dr. Stephen Moore
 Beef Genomics Laboratory
 Dept of AFNS, University of Alberta
 410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
 Tel: 780 492 0169
 Fax: 780 492 4265
 Email: stephen.moore@ualberta.ca
 The sequence best matches gb:BP019786 (Bos primigenius prothymosin
 mRNA, complete cds) in main database at high score of 735.0 and
 E-value of 0.0
 PCR Primers
 FORWARD: M13 Forward
 BACKWARD: M13 Reverse
 Seq primer: T3 primer
 High quality sequence stop: 383
 POLYA=No.

FEATURES
 source Location/Qualifiers
 1..383
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /sex="Two males and one female mixed"
 /tissue_type="Gastrointestinal tissue (GIT)"
 /cell_type="Epithelial"
 /dev_stage="Young adult"
 /lab_host="XLA-BlueMFP"-strain"
 /clone_id="Bovine Abomasum cDNA Library"
 /note="Organ: Abomasum; Vector: Uni-22APKX; Site_1: EcoR
 I; Site_2: Xho I"

BASE COUNT 83 a 111 c 113 g 76 t
 ORIGIN
 Query Match 25.6%; Score 300.6; DB 12; Length 383;
 Best Local Similarity 88.3%; Pred. No. 4e-76;
 Matches 338; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 599 AGTACTCATATCTGTTGTTGACATGATTAACGACCTAGTACTCAAGCTTGT 658
 Db 1 AGTACTCATATCTGTTGTTGACATGATTAACGACCTAGTACTCAAGCTTGT 60
 QY 659 TCTCGTTTACATGACAGATGCGCAGAGAGACATGCTCAGCTTGGAGTATTTGATC 718

Db 61 TCTCGTTTACATGACAGAGATGCGCAGAGAGACATGCTCAGCTTGGAGTATTTGATC 120
 QY 719 CATCTCACTACACAGAGATCTCTTCACTGCGGTTCCAGTACCTGTGAGAGATCGGCAAT 778
 Db 121 GGTCTTACTACACAGAGGTCCTCTGACCTGGTCCCGTGCAGTGCAGTACGATCTGCGCAGT 180
 QY 779 TCACCTGAGACAGTGTACCATCAGCGGTGTGTTGTCATGTGAAGTGGATGTCAAG 838
 Db 181 TCACCTGAGACAGTGTACCATCAGCGGTGTGTTGTCATGTGAAGTGGATGTCTCAGG 240
 QY 839 CTATCTTGATATACCGGTACGTCCAGCTGTGTGCACTTACAGCGCAGCATTTCAATTC 898
 Db 241 CCATCTCTGACACAGCGGACCTTCAAGCTGTGTGCGGCCACAGCGCAGCATCTCAACATCC 300
 QY 899 AGCAAGCTATTTGAGCCACACAGAACCAAGTACGGTGTGTTGACAT-AGATTGGACAAAC 957
 Db 301 AGCAGGCGCATTTGAGCCACACAGAACCAAGTACGGTGTGTTGACATTTGACATTTGACGACCAAC 360
 QY 958 CTTAGCTTACATGCTTACAGTTGT 980
 Db 361 CTGAGCTTACATGCTTACAGTTGT 383

RESULT 11
 B0417551/c 623 bp mRNA linear EST 23-MAY-2002
 LOCUS 1K48C09.X1 Kaestner ngm3 wt Mus musculus cDNA 3', similar to
 DEFINITION SW:CHYM SHEEP P18276 PROCHYMOSIN PRECURSOR, B-FORM ;, mRNA
 sequence.
 ACCESSION B0417551
 VERSION B0417551.1 GI:21122752
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 623)
 Lemstra,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Hillier,L., Maira,M., Page,D., Wylie,T., Martin,J., Blaisdell,A.,
 Schmitt,A., Theising,B., Ritter,B., Ronko,I., Bennett,J., Cardenas,
 M., Gibbons,M., McCann,R., Cole,R., Tesagarsishvili,R., Williams,T.,
 Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@iobp.harvard.edu
 Pancreas was obtained from Gerard Gradwohl (PNAS 97 1607-1611,
 2000) Library was constructed by Catherine Lee DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: Dr. Marie Seearce
 (msearce@mail.med.upenn.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 382.

FEATURES
 source Location/Qualifiers
 1..623
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="129/SV x CD1"
 /db_xref="taxon:10090"
 /dev_stage="p.c. 14.5"
 /lab_host="E. coli-DH12S (GIBCO)"
 /clone_id="Kaestner ngm3 wt"
 /note="Organ: pancreas; Vector: pSPORT1 (GIBCO); Site_1:
 Not I; Site_2: Sal I; The library was prepared by
 Catherine S. Lee and has not been published. The pancreas

was obtained from Gerard Gradwohl (PMAS 97 P1607-1611, 2000). The cDNAs were prepared with an oligo containing a NotI site, and SalI linkers were added to the ends. The inserts were cut with NotI before being cloned into the NotI-SalI sites in the vectors. This is one of two libraries, ngn3 wt and ngn3 -/-. The wt library is in pBOPRI, T7 promoter is 5'."

BASE COUNT 152 a 163 c 155 g 153 t
ORIGIN

Query Match 24.6%; Score 288.6; DB 13; Length 623;
Best Local Similarity 71.8%; Pred. No. 1.6e-72;
Matches 394; Conservative 0; Mismatches 149; Indels 6; Gaps 1;

QY 625 ATGATGATACCGACCTAGTACTGACAGCTTGTCTCGTTTACATGACAGGAATGGC 684
DB 623 ATGATGATACCGACCTAGTACTGACAGCTTGTCTCGTTTACATGACAGGAATGAG 564
QY 685 CAGGAGAGATGCTCAGCTTGGAGCTATGATCCATCCATGACAGGATCTCTTAC 744
DB 563 CAGGAGAGATGCTCAGCTTGGAGCTATGATCCATCCATGACAGGATCTCTTAC 504
QY 745 TGGGTTCCAGTCACTGTCAGAGTACTGCAATTCATCTGTCAGAGTCTACATCAGC 804
DB 503 TGGGTTCCAGTCACTGTCAGAGTACTGCAATTCATCTGTCAGAGTCTACATC 444
QY 805 GGTGTGTTGTTGCAATGTCAGAGTGTCAAGTATCTTGGATACCGGTACGTCAG 864
DB 443 GGTGAAGTGTGCTTGTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 384
QY 865 CTGTGTGAGCTTACGAGCTTCTCAATTCAGAGCTTATGAGGACACAGAAC 924
DB 383 TTGACGCGGCTGTGAGAGATCTCAATTCAGAGCTTATGAGGACGTCGACAGGC 324
QY 925 CAGTACGCTGATGTTGATGATGTTGCAATTCAGAGCTTATGAGGACGTCGAC 984
DB 323 CAGTACGCTGATGTTGATGATGTTGCAATTCAGAGCTTATGAGGACGTCGAC 264
QY 985 GAGATCAACGCGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1044
DB 263 GAGATCAACGCGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 204
QY 1045 TTCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104
DB 203 TTCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 150
QY 1105 TTCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1164
DB 149 TTCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 90
QY 1165 GCAATCTGA 1173
DB 89 GCAATCTGA 81

RESULT 12
CD558204 910 bp mRNA linear EST 11-JUN-2003
LOCUS AGENCOURT 14477618 NIH MGC 181 Homo sapiens cDNA clone
DEFINITION IMAGE:30355411 5', mRNA sequence.
ACCESSION CD558204
VERSION CD558204.1 GI:31584272
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 910)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics

National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDAM475 row: h column: 12
High quality sequence start: 20
High quality sequence stop: 617.
Location/Qualifiers

FEATURES

source

1..910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30355411"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_1lb="NIH MGC 181"
/note="Vector: PCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (BcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(Invitrogen). Note: this is a NIH-MGC Library."
BASE COUNT 184 a 305 c 231 g 190 t
ORIGIN

Query Match 23.2%; Score 271.8; DB 14; Length 910;
Best Local Similarity 62.8%; Pred. No. 1.6e-67;
Matches 439; Conservative 0; Mismatches 257; Indels 3; Gaps 1;

QY 245 AATACCTTGGAGAGATCTACTCGGAACCCGCTCAAGATTCACGTTCTTTGATA 304
DB 30 AGTACTTGCGACTATTCGCGATCGGAATCTCGCCAGATTTCACTGTCTTTGACA 89
QY 305 CTGTGTTCTGTGACTTCTGAGTTCTCTTATCTACTGCAAGAGCAATGCTTGACAAC 364
DB 90 CCGGCTCTTCAACTGTGGGTGCTCTGAGTCTGCTCAAGTCTGCTGCAACACC 149
QY 365 ACCAAGATTTGATCCGAGAAAGTGTCCACTTCCAGAACTTGAAGCAACCTTGTCTA 424
DB 150 ACAACCGCTTCAACCTTGAGATTTCTTCACTTCACTTCACTTCACTTCACTTCA 209
QY 425 TACACTAGGATCAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 484
DB 210 TCACTTACGCGACCGGACGATGACAGGATCTTGGATTCAGACTGTTCAGATTGGAG 269
QY 485 ACATTTGATCATTCAACAGATGATGATGATGATGATGATGATGATGATGATGAT 544
DB 270 GATCTCTGACACCAATCAGATCTTGGCTGAGGACAGAGAACTGGGCTCTTCTGT 329
QY 545 CTTATGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604
DB 330 ATTATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 389
QY 605 CGATACCTGTTGTTGCAACATGATGATGATGATGATGATGATGATGATGATGATGAT 664
DB 390 CCAACCCGCTTTGACAAATCTTGAACCGAGGCTTGTTTTCAGAGACTTCTCTG 449
QY 665 TTTACATGACAGGAATGGCCAG--GAGAGATGCTCAGCTTGAAGCTATGATTCAT 721
DB 450 TCTACTCAGCGCGGATGACAGAGTGGAGAGCTTGGATTTGAGGATTTGAGCTT 509
QY 722 CTTACTACAGAGATCTTCTCAGTGTTCAGTCACTGTGACGACGATCTGCAATTC 781
DB 510 CTTACTACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 569
QY 782 CTGTGACAGTGTCACTCAGGCGGTGTGTTTTCATGATGATGATGATGATGATGAT 841

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	570	CCGTCGACGAGCATCACCATTGAACGGAGAGGGCCATCGCTGGCGTGGAGGGCTGCCAGGCGCA	629					
Qy	842	TCTTGGATACCGGGTACGTGCACAAAGTGTGTGCAGCCTTAGCAGCGACATTTCTCAACATTGAGC	901					
Db	630	TTGTGTGACACCGGACCTCTCTGCTGACACGGGCCCAACAGCCGCCCATTTGCCAACATTCACGA	689					
Qy	902	AAGCTATTGGAGCCACACAGAACCAACGATACGGTGAATTGG	940					
Db	690	GCGACATCGGAGGCCAGCGAGAACTCAGATGCGCAGCATGG	728					
RESULT 13								
LOCUS	CD516662							
DEFINITION	AGENCOURT 14371339 NIH MGC 181 Homo sapiens cDNA clone							
VERSION	CD516662							
KEYWORDS	CD516662.1 GI:31448380							
SOURCE	EST.							
ORGANISM	Homo sapiens (human)							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.							
	1 (bases 1 to 826)							
	NIH-MGC http://mgc.nci.nih.gov/ .							
	National Institutes of Health, Mammalian Gene Collection (MGC)							
	Unpublished							
	Contact: Daniela S. Gerhard, Ph.D.							

National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: gcgabs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINTL at:
<http://image.lnl.gov>
Plate: NDMA90 row: d column: 14
High quality sequence stop: 695.

FEATURES	Location/Qualifiers
source	1. .826

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30407989"
/issue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_181"
/notes="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(invitrogen). Note: this is a NIH_MGC Library."

```

Query Match	22.9%;	Score 268.8;	DB 14;	Length 826;
Best Local Similarity	62.6%;	Pred. No. 1.1e-66;		
Matches 436; Conservative	0;	Mismatches 257;	Indels 3;	Gaps 1

QY 219 GCCACTTACCACTACTCTTGATAGTCAATCTTGGGAAGATCTACCTCGAAGCCCCGCC 278
Db 70 GCTCTGGAGAACTACTCTGGATATGAGAGTACTTGGCACTATCGGCATTCGGAACCTCTGCC 129
QY 279 TCAAGAGTCAACCCGCTCTCTTGATACAGCTGTCCTCTACTCTTCGGGTTCCTCTATCTCA 338
Db 130 CCAGGATTTCAACGTCGTCTTTGACACGGCTCTTCACACTGTGAGGTCCCTCACTCTCA 189
QY 339 CTGCAGAGCAATGCTTCGACGAACACCCAAAGATTTCAGTCCGAGAAAGTCTGCCACCTT 398

Db	190	CTGCTCCAGTCTTGGCTGACCAACACACACCGCTTAAACCTTAGAGATTCTTCCACCTA	249
OY	399	CCAGAACTTAGGCAAAACCTTGTCTTATACACTACGGTACAGGTAGCATGCAAGAACTT	458
Db	250	CCAGTCCACACGACGACAGTCTCCATCACTTACGGCACCGGACGATGACAGGATCCT	309
OY	459	AGGCTATGATACCGTCACTGTCTCAACATTGTGACATTCAACGACAGTAGACTTGG	518
Db	310	CGGATACGACACTCTCCAGGTTGAGGAGCATCTGCACCAATCAGATCTTCGGCTTG	369
OY	519	CACCAAGAAACAGAGTGAATGTCTTCACTATGACGAATTCGATGGATCTTGTATGGC	578
Db	370	CGAAGCGAACTTGACTCTTCTCTGTATTAATGCTCCCTTCGATGGCATCTTGGGCTGGC	429
OY	579	ATACCCATGCTGCGCGTCAGAGTACTCGATACCTGTGTTTGAACAATGATGAACCGCA	638
Db	430	CTACCCACGATTTCTCTCTCCGGGGCCACACCGGTCTTTGACAAATCTGGAACACAGGG	489
OY	639	CTTAGTAGCTCAAGACTTGTCTCGGTTTACATGGAACGAATGGCCAG--GAGAGCAT	695
Db	490	CTGTGTTTCTCAGGACCTCTTCTGTCTACCTCAGCCCGCATGACCAAGATGGCAGCGT	549
OY	696	GCTACGCTTGGAGCTATATGATTCATCCATACACAGAGATCTCTTCACTGGTCCAGT	755
Db	550	GGTATATCTTGTGTGGCATTGACTCTTCTTACTACACTGAAAGTCTGAAGCTGGTCCCTGT	609
OY	756	CACGTGAGCAGTACTGGCAATTCACCTGTGAGCAGTGTCAACATCACCGGTGTGTTGT	815
Db	610	TACCGTGGAGGGTTACTGGCAGATTCACCGTGGACGATCACCATGAACGAGAGGCCAT	669
OY	816	TGCATGTGAAGTGGATGTCAAGCTATCTTGGATACCGGTACGTTCAAGCTGGTGGACC	875
Db	670	CGCCTGGCGTGAAGGCTTCGACGGCCATGTGTGACACCGGCACCTCTGTGTGACCGGCC	729
OY	876	TAGCAGGCACTTCTCAACATTTACGACGAAGTAATGG	911
Db	730	AACCAAGCCCATTCACAAACATTCAGAGCGACATCGG	765

RESULT	14
LOCUS	CD499194
DEFINITION	CD499194 1101 bp mRNA linear EST 12-JUN-2003 CD438-A02.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone
ACCESSION	CD438-A02.5', mRNA sequence.
VERSION	CD499194
KEYWORDS	CD499194..1 GI:31426225
SOURCE	EST.
ORGANISM	Gasterosteus aculeatus (three spined stickleback) Gasterosteus aculeatus

REFERENCE
AUTHORS
Kingsley, D.M., Petchel, C., Balabhadra, S., Grimwood, J., Dickson, M.,

TITLE
 JOURNAL
 COMMENT
 Schmutz, J. and Myers, R.M.
 Expressed sequence tags from *Gasterosteus aculeatus*
 Unpublished
 Contact: Kingsley, DM
 HHMI and Department of Developmental Biology
 Stanford University School of Medicine
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
 Tel: 650 725 5954
 Fax: 650 725 7779
 Email: kingsley@cgm.stanford.edu
 Plate: 38
 High quality sequence. stop: 904.
 Location/Qualifiers

```

source
1. 1101
   /organism="Gasterosteus aculeatus"
   /mol_type="mrna"
   /strain="Salinas river, CA"
   /db_xref="taxon:69293"

```

```

/clone="CDA38-A02"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer). The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dt sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."
BASE COUNT      268 a      313 c      273 g      243 t      4 others
ORIGIN

```

```

Query Match      22.7%; Score 266.2; DB 14; Length 1101;
Best Local Similarity 56.7%; Pred. No. 7.6e-66;
Matches 548; Conservative 0; Mismatches 412; Indels 7; Gaps 3;

```

```

210 TGGTACGCGCCACTTACCACTTGTATGATCACTTTGGAGAGATCACTCGG 269
270 AACCCGCGCTCAAGATTTACCTTCTTTGATCTGTCTTGACCTTCTGGGTTCC 329
141 TAACTCTCTCACTCTTCACTGCTTCATCTTTGACCGGCTCTTCAACCTGAGGCTCC 200
330 CTCTATCTACTGCAAGCAATCTCTGCAAGAACCAAGATTCGATCCGAGAAATGTC 389
201 TTACAGTCTACTGTCTCAAGGAGCTTGCACCAACCAATTTCAACCCAGGAGTTC 260
390 GTCCACCTTCCAGACTTGAAGCAACCTTGTCTTATCACTAGCGATGAGATGCA 449
261 CAGCACTTTCATAGGGGAGACAGACTGTCTCATCTAGTACGAGACCGGAGCATGAC 320
450 AGGAATCTTAGGCTATGATACCTCTACTGTCTCAACATTTGAGCAATTAAAGACAGT 509
321 TGGACGCTTGGCAAGGACAAATTTGAGGAGGAGATCCGATGCCAAACAGGTTT 380
510 AGGACTTAGACCAAGCAACAGGATGATCTTCACTATGAGAAATTCAGATGGCATCT 569
381 TGGGATCAGGAGACAGA--ACCTCTTTCATGCTCAGATGGCAGCCGATGGCATCT 437
570 TGGTATGGAATCCATCGCTCGCTCAGATGACTCGATCTGTGTTTGAACAATGAT 629
438 GGGGTTGGCTTCAGAGCATGCTCTGCAATGTCGTCCTGTTTGAACAATGAT 497
630 GAACCGACACTTAGTACTAAGATTTGTTCTGGTTTATAGGACAGAAATGGCCAGGA 689
498 CCAGCAGGAGCTTGTTCACAGTCTTGTCTCCGCTTACTGAGGCGCAATAGTAGCA 557
690 GAGCA--TGCAGCGCTGGAGCTATGATCATCTACTATACAGGATCTTTCACG 746
558 GGGCAGTGAAGGTGGCTTCGGTGTATGACAGTAGCATTAACCGGTCAATACATG 617
747 GGTTCCAGTCACTGTGACAGATGAGCAATTCATCTGAGAGATGTCACATCAGCG 806
618 GATCCCTCTGAGCTCTGCGCACCTACTGGCAGATCCAAATGACAGTTCACATCATGG 677
807 TGTGTTGTTGCAATGGAAGGTGATGTCAGCTATCTTGAATACCGGTAAGTCCAACT 866
678 AAGAGCTGTGGCTGCACTGTGTCGCGCAGGCATATGACATGATGATCTCCAGAT 737
867 GGTCGAGCTAGGAGGAGCAATTCTCAACATTCAGCAAGCATATGGAGCCACAGAACCA 926
738 TGTGGCCCAACCAATGACATCAACAATGAATTCCTGTGTTGAGCCACAGCCAACCA 797

```

```

QY      927 GTACGGTGAATTGACATGATTTGGACAACTTACTACATGCTTACAGTTCTTTGA 986
DB      798 GAATGAGAGAGGCTTACAGTGAATCGCAACATCCAGGACATGCTGAAGTACCTTAC 857
QY      987 GATCAACGGCAAGATGTATCCACTGACCCCTCCGCTTATACAGCCAGGATCAAGGTT 1046
DB      858 TCTCAACGACAGAGCTTATACCATCTCTGATCTGCTTACATGATCTTACAGCTCTTACG 917
QY      1047 CTGACCACTGATGATTCAGAGTGAAGAACATTCCTCCAGAAATGATTTGGAGATGTT 1106
DB      918 CTGACACACTGGCTT-TGGCCAGGAGTGGCTCCACAGCTCTGGATCTGGAGATGCTT 976
QY      1107 CATTCGTGATCTACTACAGGCTCTTTACAGGGCCAAACCTCTGGGCTAGCTTAAAG 1166
DB      977 CATCANGCAGNTCTAGTCATCTTCCATACGAGGATCAACAAAGTTGCTGGCCCCATC 1036
QY      1167 AATCTGA 1173
DB      1037 TGTCTTA 1043

```

```

RESULT 15
CD503881      1083 bp      mRNA      linear      EST 12-JUN-2003
LOCUS      CDA64-G12.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone
DEFINITION      CDA64-G12 5', mRNA sequence.
ACCESSION      CD503881
VERSION      CD503881.1 GI:31434449
KEYWORDS      EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
ORGANISM      Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1083)

```

```

REFERENCE
AUTHORS      Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schultz,J. and Myers,R.M.
TITLE      Expressed sequence tags from Gasterosteus aculeatus
JOURNAL      Unpublished
COMMENT      Contact: Kingsley, DM
HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cwmg.stanford.edu
Plate: 64

```

```

FEATURES
source
Location/Qualifiers
1..1083
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA64-G12"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"

```

```

/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer). The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dt sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then

```

BASE COUNT 257 a 308 c 269 g 247 t 2 others
ORIGIN

Query Match 22.7%; Score 266; DB 14; Length 1083;
Best Local Similarity 56.8%; Pred. No. 8,6e-66;
Matches 549; Conservative 0; Mismatches 411; Indels 7; Gaps 3;

```
OY 210 TGCTAGCGTGCACCTTACCACTTACCTGATGTGATACCTTTGGAGAGATCTACTCGG 269
    |||
Db 46 TGGCATTTAGTTCATGACCAAGATGCCAGCTTGTCTACGATGATCTCCATCGG 105
    |||
OY 270 AACCCGCTCAAGATGATCAGCTTCTGTTGATCTGATCTGATCTTGGGTTCC 329
    |||
Db 106 TAACCTCTCTAGTCTTCAAGGTATCTTTGACACCGGCTCTCCAACTGTGGTCCC 165
    |||
OY 330 CTCTATCTACTGCAAGACATGCTGCAAGAACCAAGATTGATTCGAGAAAGTC 389
    |||
Db 166 TTCAAGTCTACTGCTCAGGAGGCTTGCAACCAATTAATTCAACCCGAGGAGTC 225
    |||
OY 390 GTCCACCTTCCAGAACTTAGGAAACCTTGTCTATACCTACGTAACAGTACATGCA 449
    |||
Db 226 CAGCACTTCCAAATGGGCGACAGACTCTGTCCATCCAGTACGACCGGACGATGAC 285
    |||
OY 450 AGGAATCTTAGGCTATGATATACGCTCATCTCTCAACATTTGGAATTCACAGACAGT 509
    |||
Db 286 TGGAGCTCTGCAAGTACATGTTGAGTGGAGGACATCAGGTGCCAAACGAGTCTT 345
    |||
OY 510 AGGACTTAGACACCAAGAACCAAGTATGTCTTCACTATGCAAGATTGATGATCCT 569
    |||
Db 346 TGGGATTCAGCGAGACAGA---AGTCCCTTTCATGGCTCAGATGGACCGAGTGCATCCT 402
    |||
OY 570 TGGTATGGCATACCATGCTGCGCTCAGAGTACTGATACCTGTGTTGACACATGAT 629
    |||
Db 403 GGGGTTGGCTTTCAGAGATGCGCTCTGACAAATGTGCGCTGTCTTTCACAAATGCT 462
    |||
OY 630 GAACCGACACCTTAGTAGCTCAAGATCTGTTCTCGGTTTACATGACAGGAATGGCCAGA 689
    |||
Db 463 CCAGAGAGGACTTGTGTTCCAGCTCTGTGTTCCGCTTACCTGAGGCGCAATATGAGCA 522
    |||
OY 690 GAGCA---TGCTCAGCTTGGAGTATGATCCATCTTACTACAGAGATCTTCACTG 746
    |||
Db 523 GGGCAGTAGAGTGGCTTCGGTGTATGACAGTAGCATTTACCGGTCAATCAGATG 582
    |||
OY 747 GGTTCAGTCACTGTGACAGCATGATGCAATTCAGTGTGACAGGTTCACCATCAGCGG 806
    |||
Db 583 GATCCCTCTGACCTGTGCCACCTACTGGCAGATCCAAATGACAGTTCACATCATGG 642
    |||
OY 807 TGTGTTGTGATGTAAGTGAAGTGCATGATCTTGTGATACCGGTACGTCAGCT 866
    |||
Db 643 ACAGGCTGTGGCTCACTGTGGCTGGCCAGGCATTCATGACATGTGTAACCTCCAGAT 702
    |||
OY 867 GGTCCGACCTAGCAGCAGCATTTCTCAACATTCAGCAAGCTATGGAGCCACAGAACCA 926
    |||
Db 703 TGTGGCCCAACCATGACATCAACAATGAAATTCGTGGGTTGAGCCACAGCCACCA 762
    |||
OY 927 GTACGTTAGTTGACATGATTTGCGACAACTTAGCTACATGCTTACAGTTGTCTTGA 986
    |||
Db 763 GAATGGAAGAGCTACAGTGAACCTGCAGAACATCCAAAGGCATGCTGAAGTCACTTCAC 822
    |||
OY 987 GATCAACGCGCAAGATGTAACCACTGACCCCTCCGCTTATACAGCCAGATCAAGGTT 1046
    |||
Db 823 TNCCTACGGAACAGGCTTACATCCCTGATCTGATGCTTCACTCAAGAGCTCTACGG 882
    |||
OY 1047 CTGCAACAGTGAATTCAGAGTGAACCAATTCAGAAATGATCTGGAGATGTGTT 1106
    |||
Db 883 CTGACGACACTGGCTT-TGGCCAGGAGGCTCCCAAGCTCTGATCTGGAGATGTCTT 941
    |||
OY 1107 CATTTGTAGTACTACAGGCTTTTGAACGGCCAAACCTGTGGCTTGGCTAGCTAAAGC 1166
    |||
Db 942 CATCAGGAGTTCATGCTATCTTGATACGGAATCAGACGTTGGCTGGGCCAGTTC 1001
    |||
```

OY 1167 AATCTGA 1173
|||
Db 1002 TGTCTAA 1008

Search completed: February 5, 2004, 05:18:27
Job time : 2899 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 04:30:48 ; Search time 4588 Seconds
(without alignments)
10459.240 Million cell updates/sec

Title: US-09-643-755B-1
Perfect score: 1173
Sequence: 1 atgaactccctaagctctt.....ggtcagctcaagcaatctga 1173

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	100.0	1173	6	AX088019
2	1173	100.0	3957	6	AX088021
3	56	4.8	1269	4	BOVCHYMOC
4	49	4.2	858	4	AF421165
5	49	4.2	957	4	AF421164
6	49	4.2	1095	4	AF421161
7	49	4.2	1098	6	E00079
8	49	4.2	1098	6	E02341
9	49	4.2	1101	4	AF177290
10	49	4.2	1115	12	SYNPROCA
11	49	4.2	1175	6	104058
12	49	4.2	1240	6	AR002347
13	49	4.2	1275	4	BOVCHYMOA
14	49	4.2	1289	6	E00144
15	49	4.2	1291	6	E15836
16	49	4.2	1305	4	BOVCHYMOB
17	49	4.2	1460	6	E00042
18	49	4.2	1460	6	E00295
19	49	4.2	2726	6	108097
20	49	4.2	2733	6	AR073077
21	49	4.2	2982	6	108098
22	49	3.8	1094	4	BP19786
23	38	3.2	843	4	AF421166
24	38	3.2	1080	4	AF421162
25	38	3.2	1290	6	E00075
26	38	3.2	1291	6	A15633
27	38	3.2	1311	6	E00108
28	37	3.2	119	4	BOVCHYMO8
29	35	3.0	171	4	BOVCHYMO2
30	35	3.0	637	6	B05472
31	35	3.0	744	4	AF421167
32	35	3.0	981	4	AF421163
33	34	2.9	1096	6	BD096860
34	33	2.8	179	6	104059
35	33	2.8	819	6	BD096861
36	33	2.8	1172	4	CDR131677
37	32	2.7	335	4	BOVCHYMO9
38	32	2.7	1292	4	CA0PCHY
39	30	2.6	134	4	BOVCHYMO6
40	30	2.6	165	4	BOVCHYMO7
41	29	2.5	1460	6	E00042
42	29	2.5	1460	6	E00295
43	27	2.3	139	4	BOVCHYMO4
44	25	2.1	138	4	BOVCHYMO3
45	23	2.0	220	4	BOVCHYMO5

ALIGNMENTS

RESULT 1
LOCUS AX088019 1173 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114571.
ACCESSION AX088019
VERSION AX088019.1 GI:13396947
KEYWORDS
SOURCE
ORGANISM
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS
TITLE
1 van Rooijen, G., Keon, R.G., Boothe, J. and Shen, Y.
Commercial production of chymosin in plants

JOURNAL Patent: WO 0114571-A 1 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES Location/Qualifiers
source 1..1173
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
CDS 1..1173
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC34680.1"
/db_xref="GI:13396948"
/translation="MNFLKSPFPYAFLCGQYFVAVTAAETRIPLVYKSLRKALK
EHLLEDFLOKQOYGISSKYSGGEVAVPLTNVLDQYFGKILYGLTPQEFTVLADT
GSSDFWVPSTYCKSNACKNORFDPKRSFTQNLGKPLSIHYGTSMQGLIDYTV
SNIVDIQQTVEGLSTOEPGVFTYAEFDGILGMAYPSLASEYSIPVDNMNRLVAD
LFSVYVDRNGOESMLTLGALDPSYTGSLHWPVTYQOYQVFTVDSVTTSGVVACEG
GCOALIDPRTSKLVGSSDILNIQALIGATONQYGFDDICDNLSTMPVTFEINRKM
YPLTPSAVYSQDGFTSGFQSENHOKWILGDVFIREFYVSFDRANMLVGLAKAI"

BASE COUNT 299 a 308 c 262 g 304 t

ORIGIN

Query Match 100.0%; Score 1173; DB 6; Length 1173;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTCTTAACTCTTCCCTTTCTACGCTTCTCTTGTTCGATCAATCTTCGTT 60
DB 1 ATGAACCTCTTAACTCTTCCCTTTCTACGCTTCTCTTGTTCGATCAATCTTCGTT 60
QY 61 GCTGTACTACGCTGTGATCAACCGGATTCCTCTCAAAAGGTAAGTCTCCGTT 120
DB 61 GCTGTACTACGCTGTGATCAACCGGATTCCTCTCAAAAGGTAAGTCTCCGTT 120
QY 121 AAGGCGCTGAAGAACATGACCTTCTAGAAAGCTTCTTGAGAAACAACAGTAGCATC 180
DB 121 AAGGCGCTGAAGAACATGACCTTCTAGAAAGCTTCTTGAGAAACAACAGTAGCATC 180
QY 181 AGCAGCAGTAGTCTCCGCTTGGTGAAGTTGCTAGCGTGCACCTTCAACCTAGTAT 240
DB 181 AGCAGCAGTAGTCTCCGCTTGGTGAAGTTGCTAGCGTGCACCTTCAACCTAGTAT 240
QY 241 AGTCATCTCTTGGGAAGATCTACCTCTGGAAACCCCGCTCAAGAGTTCAACCTCTT 300
DB 241 AGTCATCTCTTGGGAAGATCTACCTCTGGAAACCCCGCTCAAGAGTTCAACCTCTT 300
QY 301 GATACGTGTTCTCTGACTTCTGAGTTCCCTCTATCTACTGCAAGACAATGCTGCAAG 360
DB 301 GATACGTGTTCTCTGACTTCTGAGTTCCCTCTATCTACTGCAAGACAATGCTGCAAG 360
QY 361 AACCAACCAAGATTCGATCCGAGAAAGTGTCCACTTCCAGAACTTAGGCAAAACCTT 420
DB 361 AACCAACCAAGATTCGATCCGAGAAAGTGTCCACTTCCAGAACTTAGGCAAAACCTT 420
QY 421 TCTATACATGAGGTAGAGTAGATGCAAGAACTTAGGCTATGATACCTGCACTGTC 480
DB 421 TCTATACATGAGGTAGAGTAGATGCAAGAACTTAGGCTATGATACCTGCACTGTC 480
QY 481 TCCAACTTGTGACATTCACAAGACAGTAGACTTAGCAACCCAGAACAGAGTAGTGC 540
DB 481 TCCAACTTGTGACATTCACAAGACAGTAGACTTAGCAACCCAGAACAGAGTAGTGC 540
QY 541 TTCACTATGACGAATTCGATGCGATCTTGTGATGCGATACCCATGCTCGGTCAGAG 600
DB 541 TTCACTATGACGAATTCGATGCGATCTTGTGATGCGATACCCATGCTCGGTCAGAG 600
QY 601 TACTGCAATCTGTGTTGACCAATGATAAACGACCACTAGAGCTCAAGACTTGTTC 660
DB 601 TACTGCAATCTGTGTTGACCAATGATAAACGACCACTAGAGCTCAAGACTTGTTC 660
QY 661 TCGGTTTACATGACAGGAATGCGCAGAGAGATGCTCAACGCTTGAGGCTATTGATCCA 720
DB 661 TCGGTTTACATGACAGGAATGCGCAGAGAGATGCTCAACGCTTGAGGCTATTGATCCA 720

QY 721 TCTTACTACACAGGATCTCTTCACTGCGTTCCAGTCACTGTGACAGTAATCTGGCAATTC 780
DB 721 TCTTACTACACAGGATCTCTTCACTGCGTTCCAGTCACTGTGACAGTAATCTGGCAATTC 780
QY 781 ACTGTGACAGTGTCACTACGCGGTGTGTTGTCAGTGAAGTGGATGTCAAGCT 840
DB 781 ACTGTGACAGTGTCACTACGCGGTGTGTTGTCAGTGAAGTGGATGTCAAGCT 840
QY 841 ATCTTGATACCGGTAACGTCCAGCTGTGCGACCTAGACGCAATTCACATTCAG 900
DB 841 ATCTTGATACCGGTAACGTCCAGCTGTGCGACCTAGACGCAATTCACATTCAG 900
QY 901 CAAGCTATTGAGCCACACAGAACCGTACGTTGATGACATTAATGCCAACCTT 960
DB 901 CAAGCTATTGAGCCACACAGAACCGTACGTTGATGACATTAATGCCAACCTT 960
QY 961 AGCTACATGCGTCACTTGTCTTGTAGTCAACGCGAAGTGAACCACTGACCCCTCC 1020
DB 961 AGCTACATGCGTCACTTGTCTTGTAGTCAACGCGAAGTGAACCACTGACCCCTCC 1020
QY 1021 GCCTATACGACCGAGATCAAGGTTCTGCACCACTGATTCACAGTAGAACATTC 1080
DB 1021 GCCTATACGACCGAGATCAAGGTTCTGCACCACTGATTCACAGTAGAACATTC 1080
QY 1081 CAGAAATGATCTTGGAGATGTCTTCAATTCGATGACTACAGCGTCTTTGACAGGACC 1140
DB 1081 CAGAAATGATCTTGGAGATGTCTTCAATTCGATGACTACAGCGTCTTTGACAGGACC 1140
QY 1141 AACACCTGTTGGGCTAGCTAAAGCAATCTGA 1173
DB 1141 AACACCTGTTGGGCTAGCTAAAGCAATCTGA 1173

RESULT 2
AX088021
LOCUS AX088021 3957 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114571.
ACCESSION AX088021
VERSION AX088021.1 GI:13396949
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS van Rooijen, G., Keon, R.G., Boothe, J. and Shen, Y.
TITLE Commercial production of chymosin in plants
JOURNAL Patent: WO 0114571-A 3 01-MAR-2001;
Sembiosys Genetics Inc. (CA)
FEATURES Location/Qualifiers
source 1..3957
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
CDS 1554..2726
/note="Figure 2"
/codon_start=1
/transl_table=11
/protein_id="CAC34680.1"
/db_xref="GI:13396950"
/translation="MNFLKSPFPYAFLCGQYFVAVTAAETRIPLVYKSLRKALK
EHLLEDFLOKQOYGISSKYSGGEVAVPLTNVLDQYFGKILYGLTPQEFTVLADT
GSSDFWVPSTYCKSNACKNORFDPKRSFTQNLGKPLSIHYGTSMQGLIDYTV
SNIVDIQQTVEGLSTOEPGVFTYAEFDGILGMAYPSLASEYSIPVDNMNRLVAD
LFSVYVDRNGOESMLTLGALDPSYTGSLHWPVTYQOYQVFTVDSVTTSGVVACEG
GCOALIDPRTSKLVGSSDILNIQALIGATONQYGFDDICDNLSTMPVTFEINRKM
YPLTPSAVYSQDGFTSGFQSENHOKWILGDVFIREFYVSFDRANMLVGLAKAI"

BASE COUNT 1263 a 790 c 609 g 1295 t

ORIGIN

Query Match 100.0%; Score 1173; DB 6; Length 3957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTCCTTAAGTCTTTCCCTTTCACGCTTTCCTTTGGTCAATTAATCTGCT 60
Db 1554 ATGAACCTTCCTTAAGTCTTTCCCTTTCACGCTTTCCTTTGGTCAATTAATCTGCT 1613
QY 61 GCTGTATTCACGCTGCTGAGATCAACCGGCAATTCCTCTCTACAAAGTAAGTCTCCGT 120
Db 1614 GCTGTATTCACGCTGCTGAGATCAACCGGCAATTCCTCTCTACAAAGTAAGTCTCCGT 1673
QY 121 AAGGCGCTGAAGGAACATGAGCTTCTAGAACCTTCTTGAGAAACAAGATATGGCATC 180
Db 1674 AAGGCGCTGAAGGAACATGAGCTTCTAGAACCTTCTTGAGAAACAAGATATGGCATC 1733
QY 181 AGCAGCAAGTACTCCGCTGCTGAGAAAGTGTCTAGCGCACTTACCACTTACCTTGAT 240
Db 1734 AGCAGCAAGTACTCCGCTGCTGAGAAAGTGTCTAGCGCACTTACCACTTACCTTGAT 1793
QY 241 AGTCAATTCCTTGGAGAGATCTACTCGGAAACCCGCTCAAGAGTTCAACGTTCTCTT 300
Db 1794 AGTCAATTCCTTGGAGAGATCTACTCGGAAACCCGCTCAAGAGTTCAACGTTCTCTT 1853
QY 301 GATACCTGCTCTCTGATCTTCTGGGTTCCCTCTATCTATCTGCAAGAAATGCTGCAAG 360
Db 1854 GATACCTGCTCTCTGATCTTCTGGGTTCCCTCTATCTATCTGCAAGAAATGCTGCAAG 1913
QY 361 AACCAACCAAGATTCGATCCGAGAAAGTCTGACCTTCCAGAACTTAGGCAAAACCTTG 420
Db 1914 AACCAACCAAGATTCGATCCGAGAAAGTCTGACCTTCCAGAACTTAGGCAAAACCTTG 1973
QY 421 TCTATACACTACGAGTACAGTAGAGATGCAAGAAATCTTAGGTATGATACCGTCACTGTC 480
Db 1974 TCTATACACTACGAGTACAGTAGAGATGCAAGAAATCTTAGGTATGATACCGTCACTGTC 2033
QY 481 TCCAACTTGGAGACATTCACACAGACATGAGACTTACGACCCAAAGAACAGGTATGTC 540
Db 2034 TCCAACTTGGAGACATTCACACAGACATGAGACTTACGACCCAAAGAACAGGTATGTC 2093
QY 541 TTCACCTTGAAGAAATTCGATGAGCATCCTGTGTATGAGATACCCATGCTGCGCTGCAAG 600
Db 2094 TTCACCTTGAAGAAATTCGATGAGCATCCTGTGTATGAGATACCCATGCTGCGCTGCAAG 2153
QY 601 TACTGATACCTGTGTTTGAACAACATGATGAACCGACAACCTAGTACTCAAGACTTGTTC 660
Db 2154 TACTGATACCTGTGTTTGAACAACATGATGAACCGACAACCTAGTACTCAAGACTTGTTC 2213
QY 661 TCGGTTTACATGAGACAGAAATGCGCAGAGACATGCTCAAGCTTGGAGCTATGATCA 720
Db 2214 TCGGTTTACATGAGACAGAAATGCGCAGAGACATGCTCAAGCTTGGAGCTATGATCA 2273
QY 721 TCCACTACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTCAGCACTACTGCGCAATTC 780
Db 2274 TCCACTACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTCAGCACTACTGCGCAATTC 2333
QY 781 ACTGTGACAGATGTCACCATCAGCGGTGTGTTGTGATGTAAGTGAATGTCAAGCT 840
Db 2334 ACTGTGACAGATGTCACCATCAGCGGTGTGTTGTGATGTAAGTGAATGTCAAGCT 2393
QY 841 ATCTTGGATACCGGTACGTCCAGTGTGCTGCACTTACGACGACATTCCTCAACATTAG 900
Db 2394 ATCTTGGATACCGGTACGTCCAGTGTGCTGCACTTACGACGACATTCCTCAACATTAG 2453
QY 901 CAAGCTATTTGAGCCACACAGAAACAGTACGTTGATGATGATTTGGAACAACCTT 960
Db 2454 CAAGCTATTTGAGCCACACAGAAACAGTACGTTGATGATGATTTGGAACAACCTT 2513
QY 961 AGCTACATGCTTACAGTGTCTTTGAGATCAACGCGAAGATGTACCACTGACCCCTTCC 1020
Db 2514 AGCTACATGCTTACAGTGTCTTTGAGATCAACGCGAAGATGTACCACTGACCCCTTCC 2573
QY 1021 GCTTATACAGCCAGATTCAGGGTTCGACCACTGATGATTCAGAGTGAAGCAATTC 1080
Db 2574 GCTTATACAGCCAGATTCAGGGTTCGACCACTGATGATTCAGAGTGAAGCAATTC 2633

```

```

QY 1081 CAGAAATGATCTTGGAGATGTTGTTTCATTCGTGATGACTACAGCGTCTTTGACAGGCCC 1140
Db 2634 CAGAAATGATCTTGGAGATGTTGTTTCATTCGTGATGACTACAGCGTCTTTGACAGGCCC 2693
QY 1141 AACCACTGCTTGGGCTTACGTTAAAGCAATCTGA 1173
Db 2694 AACCACTGCTTGGGCTTACGTTAAAGCAATCTGA 2726

RESULT 3
BOVCHYMOC
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Original source text: Bos taurus calf fourth stomach mucosa cDNA to
mRNA.
chymosin is the major proteolytic enzyme in the fourth stomach of
the unweaned calf. two chromatographically different forms, a and
b, of the enzyme and its precursor are known and a third form,
represented by this sequence, seems likely. the presence of a
termination codon (bases 27-29) within the prepropeptide coding
sequence implies that this sequence might be a pseudogene mtna. or
there may be an error in the cdna. the amino acid sequence deduced
from this nucleotide sequence differs at three sites from bovine
chymosin a (see bovchymoa) and at four sites from bovine chymosin b
(see bovchymob).

FEATURES
source
Location/Qualifiers
1..1269
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Fourth stomach mucosa"
/dev_stage="calf"
<1..1269
/product="chymoc mRNA"
402..1130
/note="precursor"
/codon_start=1
/product="chymosin C"
/protein_id="AAA30449.1"
/db_xref="GI:457097"
/translation="MOSGILGVDPYTVSNIVDIOOTVGLSTQEPDVFVTAEPDILICM
ATPSLASRYSIPVFDNMAMRHVLAQDIFSYMDNCESEMLTGLADIPSYTSGSHVY
PVTVOQDQWQFTVDSVITSGVAVACGGCOALIDTGISLVGPSSDIINIOQATIGATON
QYDFEIDQDNLSTYMPVFEINGKIVLPISAYTGODDGFCTTGFSEHNSQKMIIG
DVFFREYYSVFDNRANIVGLAKTI"

BASE COUNT
299 a 381 c 321 g 268 t

ORIGIN
Query Match 4.8%; Score 56; DB 4; Length 1269;
Best local Similarity 100.0%; Pred. No. 7.3e-18;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 GTGAGAGATGATGCGCAATTCATGTGAGACAGTCAACATCAGCGGTGTGTGT 815
Db 717 GTGAGAGATGATGCGCAATTCATGTGAGACAGTCAACATCAGCGGTGTGTGT 772

RESULT 4

```

AF421165 858 bp mRNA linear MAM 01-APR-2002
LOCUS Bos taurus chymosin precursor, mRNA, partial cds, alternatively
spliced.
DEFINITION AF421165
ACCESSION AF421165 GI:19851897
VERSION
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
TITLE Multiple splicing forms of bovine prochymosin mRNA, their
identification and characterization
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 858)
Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
REFERENCE Direct Submission
AUTHORS Submitted (19-SEP-2001) Institute of Animal Breeding and Genetics,
JOURNAL location/Qualifiers
TITLE VMU, Veterinaerplatz 1, Vienna 1210, Austria
FEATURES
source
1. .858
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="fourth stomach mucosa"
/dev_stage="calf"
/EC_number="3.4.23.4"
/note="rennin; preproprotein; alternatively spliced; lacks
exons 3, 4 and 8"
/codon_start=2
/product="chymosin precursor"
/protein_id="AAL99910.1"
/db_xref="GI:19851898"
/translation="VLLAVFALSGTEITRIPIYKGSIRKALKEHGLIEDFLQKOY
GISKTSIGREAVSPPLTYLDVSNIVDIQQTGSLTQEGDVFYTAEPGILGMAYP
SLASEYSIPVFDNMNRHLVAQDLPSVYMDRNGQESMLTLGALIDPSYTGSLHWVPT
VOOYQFIVDSYTIISGVVACGCGCALIDTGTSLKLVGPSDIINIQALIGATONQYG
EDGFCSTGFSQSNHSQKWLIGVFIREFYVSVPFRANNVLGLAKAI"
BASE COUNT 193 a 255 c 241 g 169 t
ORIGIN
Query Match 4.2%; Score 49; DB 4; Length 858;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 656 TGTTCGCGTTTACATGACGAGAAATGCCAGAGACATGCTCAGCCT 704
Db 378 TGTTCGCGTTTACATGACGAGAAATGCCAGAGACATGCTCAGCCT 426
RESULT 5
AF421164 957 bp mRNA linear MAM 01-APR-2002
LOCUS Bos taurus chymosin precursor, mRNA, partial cds, alternatively
spliced.
DEFINITION AF421164
ACCESSION AF421164 GI:19851895
VERSION
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
TITLE Multiple splicing forms of bovine prochymosin mRNA, their
identification and characterization
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 957)
Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
REFERENCE Direct Submission
AUTHORS Submitted (19-SEP-2001) Institute of Animal Breeding and Genetics,
JOURNAL location/Qualifiers
TITLE VMU, Veterinaerplatz 1, Vienna 1210, Austria
FEATURES
source
1. .957
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="fourth stomach mucosa"
/dev_stage="calf"
/EC_number="3.4.23.4"
/note="rennin; preproprotein; alternatively spliced; lacks
exons 3 and 4"
/codon_start=2
/product="chymosin precursor"
/protein_id="AAL99909.1"
/db_xref="GI:19851896"
/translation="VLLAVFALSGTEITRIPIYKGSIRKALKEHGLIEDFLQKOY
GISKTSIGREAVSPPLTYLDVSNIVDIQQTGSLTQEGDVFYTAEPGILGMAYP
SLASEYSIPVFDNMNRHLVAQDLPSVYMDRNGQESMLTLGALIDPSYTGSLHWVPT
VOOYQFIVDSYTIISGVVACGCGCALIDTGTSLKLVGPSDIINIQALIGATONQYG
EDIDCDNLSTYMPVTFEINGKMPILTPSAVTSOPQGFCTSGFQSNHSQKWLIGBVF
IREYVSVPFRANNVLGLAKAI"

AUTHORS Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Institute of Animal Breeding and Genetics,
VMU, Veterinaerplatz 1, Vienna 1210, Austria
FEATURES
source
1. .957
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="fourth stomach mucosa"
/dev_stage="calf"
/EC_number="3.4.23.4"
/note="rennin; preproprotein; alternatively spliced; lacks
exons 3 and 4"
/codon_start=2
/product="chymosin precursor"
/protein_id="AAL99909.1"
/db_xref="GI:19851896"
/translation="VLLAVFALSGTEITRIPIYKGSIRKALKEHGLIEDFLQKOY
GISKTSIGREAVSPPLTYLDVSNIVDIQQTGSLTQEGDVFYTAEPGILGMAYP
SLASEYSIPVFDNMNRHLVAQDLPSVYMDRNGQESMLTLGALIDPSYTGSLHWVPT
VOOYQFIVDSYTIISGVVACGCGCALIDTGTSLKLVGPSDIINIQALIGATONQYG
EDIDCDNLSTYMPVTFEINGKMPILTPSAVTSOPQGFCTSGFQSNHSQKWLIGBVF
IREYVSVPFRANNVLGLAKAI"
BASE COUNT 218 a 288 c 260 g 191 t
ORIGIN
Query Match 4.2%; Score 49; DB 4; Length 957;
Best Local Similarity 100.0%; Pred. No. 3.9e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 656 TGTTCGCGTTTACATGACGAGAAATGCCAGAGACATGCTCAGCCT 704
Db 378 TGTTCGCGTTTACATGACGAGAAATGCCAGAGACATGCTCAGCCT 426
RESULT 6
AF421161 1095 bp mRNA linear MAM 01-APR-2002
LOCUS Bos taurus chymosin precursor, mRNA, partial cds, alternatively
spliced.
DEFINITION AF421161
ACCESSION AF421161 GI:19851889
VERSION
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
TITLE Multiple splicing forms of bovine prochymosin mRNA, their
identification and characterization
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1095)
Zinovieva, N., McCoy, B., Mueller, M. and Brem, G.
REFERENCE Direct Submission
AUTHORS Submitted (19-SEP-2001) Institute of Animal Breeding and Genetics,
JOURNAL location/Qualifiers
TITLE VMU, Veterinaerplatz 1, Vienna 1210, Austria
FEATURES
source
1. .1095
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="fourth stomach mucosa"
/dev_stage="calf"
/EC_number="3.4.23.4"
/note="rennin; preproprotein; alternatively spliced; lacks
exon 8"
/codon_start=2
/product="chymosin precursor"

/protein_id="AAL99906.1"
/db_xref="GI:19851890"
/translation="VLAIVFALSCQTEITRIPLYKSKLRKALKEHGLEPFIKQOY
GISKXSGFGEVAVSPITNVLDSOYFEGKIYGTPOEFTVLETPSSPFWPSTYCKS
NGCKKHOFDPKRSSTPONIGKPLSIHGTGSMOGLIGVDITVSNIVDIOCTGLST
QEPGVFTYAEFDGILGMAYPSIASEYSIPFEDNMNHNHVAQDI.FSYIMDRNGESM
LTLGALDPSYTGSLHWVPTVQOQWOTVDSVTIISGVVACGGCAQLDGTSLV
GPSSDILNIQQAIGATQWQYGBDQFCFCTSGSEHNSQKMLIGDVFIREYYSVFDNRAN
"NLVGLAKAI"

BASE COUNT 246 a 333 c 298 g 218 t

ORIGIN

Query Match 4.2%; Score 49; DB 4; Length 1095;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCT 704
|||||
615 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCT 663

Db

RESULT 7
E00079 1098 bp RNA linear PAT 29-SEP-1997
LOCUS E00079
DEFINITION cDNA sequence encoding calf prorennin.
ACCESSION E00079
VERSION E00079.1 GI:2168383
KEYWORDS JP 1983032896-A/1.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1098)
AUTHORS Beppu, T., Uozumi, T. and Nishimori, K.
TITLE CONJUGATED PLASMOD AND MICROORGANISM CONTAINING THE SAME
JOURNAL Patent: JP 1983032896-A 1 25-FEB-1983;
BEPPI TERUHIKO
OS calf
PN JP 1983032896-A/1
PD 25-FEB-1983
PF 24-AUG-1981 JP 1981131631
PI BEPPI TERUHIKO, UOZUMI TAKESHI, NISHIMORI KATSUHIKO PC
C07H21/04, C12N1/00, C12N15/00, C12N1/00, C12R1:19; CC
strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=mucous membrane of forth stomach; CC
*source: clone=PCR 100 1;
FH Key Location/Qualifiers
FH CDS 1..1098
FT /product='calf prorennin',
LOCATION/Qualifiers
1..1098
/organism="unidentified"
/mol_type="genomic RNA"
/db_xref="taxon:32644"

BASE COUNT 252 a 327 c 300 g 219 t

ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1098;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCT 704
|||||
581 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCT 629

Db

RESULT 8
E02341 1098 bp RNA linear PAT 29-SEP-1997
LOCUS E02341
DEFINITION cDNA sequence coding for infant bovine prorennin.

ACCESSION E02341
VERSION E02341.1 GI:2170576
KEYWORDS JP 1990109984-A/1.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bubalus.
1 (bases 1 to 1098)
Beppu, T., Uozumi, T. and Nishimori, K.
TITLE COMPLEX PLASMOD AND MICROORGANISM CONTAINING THE SAME
JOURNAL Patent: JP 1990109984-A 1 23-APR-1990;
BEPPI TERUHIKO
OS Bovine
PN JP 1990109984-A/1
PD 23-APR-1990
PF 01-DEC-1988 JP 198802176
PI BEPPI TERUHIKO, UOZUMI TAKESHI, NISHIMORI KATSUHIKO PC
C12N15/59, C12N1/21;
strandedness: Single;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=Stomach mucosa;
FH Key Location/Qualifiers
FH CDS 1..1098
FT /product='Infant bovine prorennin',
LOCATION/Qualifiers
1..1098
/organism="Bos taurus"
/mol_type="genomic RNA"
/db_xref="taxon:9913"

BASE COUNT 252 a 327 c 300 g 219 t

ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1098;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCT 704
|||||
581 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCT 629

Db

RESULT 9
AF177290 1101 bp mRNA linear MAM 24-JAN-2000
LOCUS AF177290
DEFINITION Bubalus arnee bubalis prothymosin mRNA, complete cds.
ACCESSION AF177290
VERSION AF177290.1 GI:6739579
KEYWORDS
SOURCE
ORGANISM Bubalus bubalis (water buffalo)
Bubalus bubalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bubalus.
1 (bases 1 to 1101)
Bacish, V.K., Mukhopadhyay, U.K., Mohanty, A.K., Grover, S. and
Kuijpers, O.P.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1999) Molecular Biology Unit, Animal
Biotechnology Centre, National Dairy Research Institute, G.T.Road,
Karnal, Haryana 132001, India
LOCATION/Qualifiers
1..1101
/organism="Bubalus bubalis"
/mol_type="mRNA"
/sub_species="bubalis"
/db_xref="taxon:89462"
1..1101
/codon_start=1
/product="prothymosin"

```

/protein_id="AAF27315.1"
/db_xref="GI:6739580"
/translation="MAETRIPLCKSKSLRKALKENGLEDFLOKQOYVSSKYSQFG
EVAASVPLTVLDSQYFGKTLGTPEQETVLPTGSSDPWBSITCKSNACKNHRFD
PKRSSTFONIGKPLSIRYGTSMQGLIGDTATVSNIVIQVGLSTQEPDVFYTA
EPDGLIGMAVPSLASEYSIPVFDNMNRHLVADLPSVMDRGRSMTLTGALIPSY
YTGSLHWVTVQOQFTVDSITISGVAVACGGCOALIDGTSLVGPSSDILNIO
CAIGATONOGYERDIDCDNLASWPTVSVSLNGLMYLTSAIYSQDQGCISFGQSEN
RSGQMLGCVFIHREIYSVFDNRANLVGLAKAT"

BASE COUNT      244 a      336 c      306 g      215 t
ORIGIN

Query Match      4.2%; Score 49; DB 4; Length 1101;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      656 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 704
      |||||||
Db      584 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 632

RESULT 10
LOCUS      SYNPROCAA      1115 bp      DNA      linear      SYN 27-APR-1993
DEFINITION      Synthetic bovine prochymosin A protein region.
ACCESSION      M22593 M18758
VERSION      M22593.1 GI:209139
KEYWORDS      prochymosin.
SOURCE      synthetic construct
ORGANISM      synthetic construct
REFERENCE      1 (bases 1 to 1115)
AUTHORS      Wosniak,M.A., Barnett,R.W., Vicentini,A.M., Erfle,H., Elliott,R.,
      Sumner-Smith,M., Mantel,N. and Davies,R.W.
      Rapid construction of large synthetic genes: total chemical
      synthesis of two different versions of the bovine prochymosin gene
JOURNAL      Gene 60 (1), 115-127 (1987)
MEDLINE      88152494
PUBMED      3126097
COMMENT      Original source text: Synthetic DNA.
FEATURES
      source
      location/Qualifiers
      1..1115
      /organism="synthetic construct"
      /mol_type="genomic DNA"
      /db_xref="taxon:32630"
      /db_xref="taxon:32630"
      227 t
BASE COUNT      254 a      332 c      302 g      227 t
ORIGIN      1 bp downstream of HindIII site.

Query Match      4.2%; Score 49; DB 12; Length 1115;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      656 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 704
      |||||||
Db      580 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 628

RESULT 11
LOCUS      104058      1175 bp      DNA      linear      PAT 02-DEC-1994
DEFINITION      Sequence 5 from Patent EP 0123928.
ACCESSION      104058
VERSION      104058.1 GI:591912
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
COMMENT      Unclassified.
      1 (bases 1 to 1175)
REFERENCE      Cashion,L.M., McCaman,M.T., Rice,C.W. and Siae,S.R.
AUTHORS      Recombinant DNA coding for a polypeptide displaying milk clotting
      activity
JOURNAL      Patent: EP 0123928-A2 5 07-NOV-1984;
FEATURES      Location/Qualifiers

```

```

source      1..1175
/organism="unknown"

BASE COUNT      263 a      356 c      318 g      238 t
ORIGIN

Query Match      4.2%; Score 49; DB 6; Length 1175;
Best Local Similarity 100.0%; Pred. No. 3.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      656 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 704
      |||||||
Db      649 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 697

RESULT 12
LOCUS      AR002347      1240 bp      DNA      linear      PAT 04-DEC-1998
DEFINITION      Sequence 2 from patent US 5741665.
ACCESSION      AR002347
VERSION      AR002347.1 GI:3963901
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      Unclassified.
      1 (bases 1 to 1240)
AUTHORS      Kato,E.K. and Stuart,W.Dorsey.
      Light-regulated promoters for production of heterologous proteins
      in filamentous fungi
JOURNAL      Patent: US 5741665-A 2 21-APR-1998;
FEATURES      Location/Qualifiers
      source
      1..1240
      /organism="unknown"

BASE COUNT      274 a      374 c      339 g      253 t
ORIGIN

Query Match      4.2%; Score 49; DB 6; Length 1240;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      656 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 704
      |||||||
Db      700 TGTTCCTCGTTTACATGACAGCAATGGCCAGAGAGAGCATGCTCACGCT 748

RESULT 13
LOCUS      BOVCHYM0A      1275 bp      mRNA      linear      MAM 19-DEC-2002
DEFINITION      Bos taurus preprochymosin a mRNA, complete cds.
ACCESSION      J00002
VERSION      J00002.1 GI:162857
KEYWORDS
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
      Bovidae; Bovinae; Bos.
      1 (bases 1 to 1275)
AUTHORS      Moir,D., Mao,J., Schumm,J.W., Vovis,G.F., Alford,B.L. and
      Taunton-Rigby,A.
      Molecular cloning and characterization of double-stranded cDNA
      coding for bovine chymosin
JOURNAL      Gene 19 (1), 127-138 (1982)
MEDLINE      83054629
PUBMED      6183168
COMMENT      chymosin (rennin) is the major proteolytic enzyme in the fourth
      stomach of the unweaned calf. two chromatographically distinct
      forms are known and a third seems likely (see bovchymob and
      bovchymoc). this sequence has been tentatively identified as
      chymosin a; it differs from chymosin b by only two amino acids and
      from chymosin c by another amino acid. [1] argues that the
      different chymosins are probably polymorphic variants of a single
      chymosin gene.
FEATURES      Location/Qualifiers

```

source 1..1275
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="fourth stomach mucosa"
21..1166
/note="femlin"
/codon_start=1
/product="preprochymosin a"
/protein_id="AAA30447.1"
/db_xref="GI:162858"
/translation="MRCLVLLAVPALVSLQGAETIRPIYKGSLSKALKKEHLEDEL
QKQYGISSTKSGEGLVPLTNLYDSQYFGKTYLGPQEFVLFDPTGSSFWVPS
LYCKSNACKNQRPDPKRSSTFONLGRPLSIHYGSGQILGYDTVTVSNILDIQT
VGLSTQEPDVFTYAEFDGILGMAYPSLASEYSIPVPMNMNRHLVAODLFSVMDRN
GOESMLTGLAIDPCYTGSLHWVPTVOQYQVTDVSTISGVAVACGGCOALIDTG
TSKLVGPSDLINIQALGATQNGYDERDIDCDNISYPTVVFELNGMYPLTTSAYT
SDQGFCTSGFQSEHNSQKWLGDVFIKEYTSVFDNRANNVGLAKAT"
24..68
/note="presequence"
mat_peptide 195..1163
/product="chymosin a"
BASE COUNT 293 a 391 c 336 g 255 t
ORIGIN

Query Match 4.2%; Score 49; DB 4; Length 1275;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGAAATGGCCAGAGAGCATGCTCAGCT 704
|||||
649 TGTTCGCGTTTACATGACAGAAATGGCCAGAGAGCATGCTCAGCT 697
|||||

RESULT 14
E00144 1289 bp RNA linear PAT 29-SEP-1997
LOCUS CDNA encoding bovine calf chymosin.
DEFINITION E00144
ACCESSION E00144
VERSION E00144.1 GI:2168443
KEYWORDS JP 1984021392-A/1.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1289)
AUTHORS Chiyaarusu, B.B.
TITLE CATTLE KIMOCIN
JOURNAL Patent: JP 1984021392-A 1 03-FEB-1984;
GENEX CORP
OS bovine calf
PN JP 1984021392-A/1
PD 03-FEB-1984
PF 30-JUN-1983 JP 1983119481
PR 01-JUL-1982 US 82 394433, 13-APR-1983 US 83 484539 PI
CHITAAARUSU EI BASURETSUTO
PC C12N15/00, C07H21/04, C12N1/20, C12P19/34, C12P21/02, (C12N15/00,
PC C12R1:19);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue_type=stomach mucosa;
FH key
FH Location/Qualifiers
FT mat_peptide 169..1113
FT /product='bovine calf chymosin' FT CDS
FT <1..1116
FT /product='precursor polypeptide'.
FEATURES
source 1..1289
Location/Qualifiers
1..1289
/organism="unclassified"
/mol_type="genomic RNA"
/db_xref="taxon:32644"

BASE COUNT 311 a 388 c 329 g 260 t 1 others
ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1289;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGAAATGGCCAGAGAGCATGCTCAGCT 704
|||||
623 TGTTCGCGTTTACATGACAGAAATGGCCAGAGAGCATGCTCAGCT 671
|||||

RESULT 15
A15836 1291 bp DNA linear PAT 16-FEB-1994
LOCUS chymosin gene.
DEFINITION A15836
ACCESSION A15836
VERSION A15836.1 GI:488959
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Simons, A.F.M., and De Vos, W.M.
TITLE DNA fragments, containing a lactic acid bacterium-specific
regulator region for the expression of genes coding for normally
heterologous proteins
Patent: EP 0307011-A 5 15-MAR-1989;
NEDERLANDS INSTITUUT VOOR ZUIVELANDERBOEK
FEATURES
source 1..1291
Location/Qualifiers
1..1291
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
BASE COUNT 323 a 382 c 328 g 257 t 1 others
ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGAAATGGCCAGAGAGCATGCTCAGCT 704
|||||
609 TGTTCGCGTTTACATGACAGAAATGGCCAGAGAGCATGCTCAGCT 657
|||||

Search completed: February 5, 2004, 07:24:19
Job time : 4594 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 03:13:23 ; Search time 390 Seconds
(without alignments)
8119.079 Million cell updates/sec

Title: US-09-643-755B-1
Perfect score: 1173
Sequence: 1 atgaacttccttaagtcctt.....gctagctaaagcaatctga 1173

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: N_Geneseq_19Jun03.*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 26: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	1173	100.0	1173	22	AA500569
2	1173	100.0	3957	22	AA500570
3	49	4.2	1098	4	AA330063
4	49	4.2	1098	11	AA004683
5	49	4.2	1175	5	AA40295
6	49	4.2	1175	13	AA020949
7	49	4.2	1210	12	AA014051
8	49	4.2	1240	16	AA03006

9	49	4.2	1278	5	AA40055
10	49	4.2	1289	4	AA330022
11	49	4.2	1291	10	AA291157
12	49	4.2	1314	4	AA330049
13	49	4.2	1460	3	AA20043
14	49	4.2	1460	5	AA40180
15	49	4.2	1460	3	AA40180
16	49	4.2	2727	10	AA291188
17	49	4.2	2733	20	AA206463
18	49	4.2	2733	24	AA53073
19	49	4.2	2982	10	AA291185
20	38	3.2	1290	4	AA330209
21	35	3.0	637	14	AA049459
22	34	2.9	1096	20	AA40296
23	33	2.8	179	5	AA40296
24	33	2.8	186	13	AA020950
25	32	2.8	819	20	AA402950
26	32	2.7	107	12	AA014777
27	29	2.5	1460	3	AA20043
28	29	2.5	1460	5	AA40180
29	23	2.0	609	24	AB06309
30	23	2.0	1083	21	AA044658
31	23	2.0	1140	24	AA597159
32	23	2.0	1238	24	AB58369
33	23	2.0	1521	24	AB214440
34	20	1.7	1814	21	AA037875
35	20	1.7	828	11	AA003224
36	20	1.7	846	11	AA003223
37	20	1.7	1104	9	AA080001
38	20	1.7	1143	5	AA40214
39	20	1.7	1146	24	AB553735
40	19	1.6	116624	19	AAV52850
41	19	1.6	231	23	AA55374
42	19	1.6	556	24	ABV87911
43	19	1.6	689	19	AAV59757
44	19	1.6	689	24	AB573750
45	19	1.6	1353	21	AA239862
			1751	19	AAV59610

ALIGNMENTS

RESULT 1	AA500569	standard; DNA; 1173 BP.
ID	AA500569	
AC	AA500569;	
XX		
DT	14-MAY-2001	(first entry)
DE		Bovine pre-pro-chymosin DNA sequence.
XX		
KW		Chymosin; transcription regulator; terminator sequence; soybean; corn;
KW		pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;
KW		barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;
KW		safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;
KW		squash; jojoba; ds.
OS		Bos sp.
XX		
XX		
Key		Location/Qualifiers
CDS		1..1173
FT		/*tag= a
FT		/product= "Bovine chymosin"
FT		sig_peptide
FT		1..78
FT		/*tag= b
FT		/product= "Bovine chymosin"
FT		misc_feature
FT		79..201
FT		/*tag= c
FT		/note= "Pro sequence"
FT		mat_peptide
FT		202..1170
FT		/*tag= d
FT		/product= "Mature bovine chymosin"
XX		

Sequence of prochy
Sequence of veal c
Cloned sequence of
cDNA sequence corr
Pre-prorennin-A ge
Sequence of recomb
BamHI/SalI insert
2.7 Kbp HindIII fr
DNA encoding oleos
BamHI insert from
Sequence of prepro
Prochymosin gene 5
DNA encoding the f
Sequence coding fo
DNA coding for a r
DNA encoding His-c
Remnin - casein co
Pre-prorennin-A ge
Sequence of recomb
Arabidopsis thalia
Arabidopsis thalia
Human aspartyl pro
Protein modificati
Arabidopsis thalia
Arabidopsis thalia
DNA fragment of PA
DNA fragment of PA
Optimised prochymo
Gene encoding (pre
Aspergillus gene e
Human eyal gene co
cDNA #50 encoding
Human secreted pro
Human cDNA #2 for
Plasmodium ovale p
Human secreted pro

PN MO200114571-A1.
 XX
 PD 01-MAR-2001.
 XX
 PF 23-AUG-2000; 2000MO-CA00975.
 XX
 PR 23-AUG-1999; 99US-0378696.
 XX
 PA (SEMB-) SEMBIOSIS GENETICS INC.
 XX
 PI Van Rooijen G, Keon RG, Boothe J, Shen Y;
 XX
 DR WPI; 2001-226621/23.
 XX P-PSDB; AAU00536.
 XX
 PT Producing chymosin in seeds of plants such as rice, flax, rape seed, by
 XX transforming plant cell with a nucleic acid encoding chymosin operably
 XX linked to transcription regulator and terminator sequences -
 PS Claim 9; Fig 1; 56pp; English.

CC The sequence represents a DNA which encodes a bovine chymosin
 CC polypeptide. Chymosin can be produced in a plant seed through
 CC introduction of a chimeric nucleic acid molecule, comprising a nucleic
 CC acid sequence encoding a chymosin polypeptide operatively linked to
 CC transcription regulator and terminator sequences, into a plant cell. The
 CC sequences are useful for producing plant seeds, in particular seeds of
 CC soybean, rape seed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
 CC barley, oats, sorghum, Arabidopsis thaliana, potato, flax/linseed,
 CC safflower, oil palm, groundnut, Brazil nut, coconut, castor, coriander,
 CC squash, jojoba and rice.

Sequence 1173 BP; 299 A; 308 C; 262 G; 304 T; 0 other;

Query Match 100.0%; Score 1173; DB 22; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTCTTAACTCTTCCCTTTCTAGCGCTTCTCTTGGTCAATACCTCGTT 60
 DB 1 ATGAACCTCTTAACTCTTCCCTTTCTAGCGCTTCTCTTGGTCAATACCTCGTT 60
 QY 61 GCTGTTACTACGCTGCTGAGATCAGCCGATTCCTCTCAAAAGTAAGTCTCTCGT 120
 DB 61 GCTGTTACTACGCTGCTGAGATCAGCCGATTCCTCTCAAAAGTAAGTCTCTCGT 120
 QY 121 AAGCGCTGAAGAAACATGACCTTCTAGAAAGCTTCTTGAGAAACAACAGTAGCATC 180
 DB 121 AAGCGCTGAAGAAACATGACCTTCTAGAAAGCTTCTTGAGAAACAACAGTAGCATC 180
 QY 181 AGCAGCAGTAGTCTCCGCTTCTGTAAGTGTCTAGCGTGCCTTACCACTACCTTGAT 240
 DB 181 AGCAGCAGTAGTCTCCGCTTCTGTAAGTGTCTAGCGTGCCTTACCACTACCTTGAT 240
 QY 241 AGTCAATCTTGGAGATCTACCTGGAAACCCCGCTCAAGGTTCAACCTTCTCTT 300
 DB 241 AGTCAATCTTGGAGATCTACCTGGAAACCCCGCTCAAGGTTCAACCTTCTCTT 300
 QY 301 GATACCTGCTCTGATCTGAGTCTGCGTCTCTCTTACTGCAAGCAATGCTGCAAG 360
 DB 301 GATACCTGCTCTGATCTGAGTCTGCGTCTCTCTTACTGCAAGCAATGCTGCAAG 360
 QY 361 AACCAACAAAGATTGATCCGAGAAAGTCTGCACTTCCAGAACTTAGGCAAACTCTTG 420
 DB 361 AACCAACAAAGATTGATCCGAGAAAGTCTGCACTTCCAGAACTTAGGCAAACTCTTG 420
 QY 421 TCTATACCTACCGTATAGTAGTATGCAAGAACTTAGGCTATGATCCGTACCTGTC 480
 DB 421 TCTATACCTACCGTATAGTAGTATGCAAGAACTTAGGCTATGATCCGTACCTGTC 480
 QY 481 TCCAACTTGTGATGATTCAGACAGACAGTGAAGCTTGAAGCAAGCAAGCAAGTATGTC 540
 DB 481 TCCAACTTGTGATGATTCAGACAGACAGTGAAGCTTGAAGCAAGCAAGCAAGTATGTC 540

QY 541 TTCACCTATGCAAGATTGATGATCTCTTGGTATGAGCATACCCATCGCTCGGCTGAG 600
 DB 541 TTCACCTATGCAAGATTGATGATCTCTTGGTATGAGCATACCCATCGCTCGGCTGAG 600
 QY 601 TACTGATACCTGTTGTTGACAAATGATGAACCGACACCTTAGTACTGAAGCTTGTTC 660
 DB 601 TACTGATACCTGTTGTTGACAAATGATGAACCGACACCTTAGTACTGAAGCTTGTTC 660
 QY 661 TCGGTTTACATGAGCAAGAAATGCGCAGAGAGATGCTCAGCGTTGAGCTATGATCCA 720
 DB 661 TCGGTTTACATGAGCAAGAAATGCGCAGAGAGATGCTCAGCGTTGAGCTATGATCCA 720
 QY 721 TCTTACTACAGAGATCTCTTCACTGAGTTCCAGTACTGTGACAGATCTGAGCAATTC 780
 DB 721 TCTTACTACAGAGATCTCTTCACTGAGTTCCAGTACTGTGACAGATCTGAGCAATTC 780
 QY 781 ACTGTGACAGTGTCCACATCAGCGGTGTTGTTGATGATGAAGTGAATGTCAGCT 840
 DB 781 ACTGTGACAGTGTCCACATCAGCGGTGTTGTTGATGATGAAGTGAATGTCAGCT 840
 QY 841 ATCTTGATACCGGTACGTCACAGCTGTGCGACCTAGCAGGCAATTCATTCAG 900
 DB 841 ATCTTGATACCGGTACGTCACAGCTGTGCGACCTAGCAGGCAATTCATTCAG 900
 QY 901 CAAGCTATTTGAGCCCAAGAAACCGATACGTTGATGATGATGATGATGATGATGAT 960
 DB 901 CAAGCTATTTGAGCCCAAGAAACCGATACGTTGATGATGATGATGATGATGATGAT 960
 QY 961 AGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 961 AGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 QY 1021 GCTTATACAGCAGGATCAAGGTTCTGACAGTGTGATGATGATGATGATGATGATGAT 1080
 DB 1021 GCTTATACAGCAGGATCAAGGTTCTGACAGTGTGATGATGATGATGATGATGATGAT 1080
 QY 1081 CAGAAATGATCTTGGAGATGTTCTTCTGATGATGATGATGATGATGATGATGATGAT 1140
 DB 1081 CAGAAATGATCTTGGAGATGTTCTTCTGATGATGATGATGATGATGATGATGATGAT 1140
 QY 1141 AACCACTGTTGGGCTAGCTAAAGCAATCTGA 1173
 DB 1141 AACCACTGTTGGGCTAGCTAAAGCAATCTGA 1173

RESULT 2
 ID AAS00570 standard; DNA: 3957 BP.
 XX AAS00570;
 AC
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.
 XX
 XX Chymosin; transcription regulator; terminator sequence; soybean; corn;
 KW pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;
 KW barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;
 KW safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;
 KW squash; jojoba; de; phaseolin; promoter; mutant; French bean.
 XX
 OS Chimeric - Bos sp.
 XX Chimeric - Phaseolus vulgaris.
 OS
 FH Key Location/Qualifiers
 FT 1..1553
 FT /*tag= a
 FT /note= "Phaseolin promoter"
 FT 1554..2726
 FT /*tag= b
 FT /product= "Bovine pre-pro-chymosin"
 FT Terminator 2727..3957

```

FT      /*tag= C
PN      /note= "Phaseolin terminator"
XX      MO200114571-AI.
XX      01-MAR-2001.
XX      23-AUG-2000; 2000MO-CA00975.
XX      23-AUG-1999; 99US-0378696.
XX      (SEMB-) SEMBIOSYS GENETICS INC.
XX      Van Rooijen G, Keon RG, Boothe J, Shen Y;
XX      WPI; 2001-226621/23.
XX      P-PSDB; AAU00536.
XX      Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT      transforming plant cell with a nucleic acid encoding chymosin operably
PT      linked to transcription regulator and terminator sequences -
XX      Example 1; Fig 2; 56pp; English.
XX      The sequence represents a chimeric polynucleotide comprising a
CC      pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
CC      Chymosin can be produced in a plant seed through introduction of a
CC      chimeric nucleic acid molecule, comprising a nucleic acid sequence
CC      encoding a chymosin polypeptide operatively linked to transcription
CC      regulator and terminator sequences, into a plant cell. The sequences are
CC      useful for producing plant seeds, in particular seeds of soybean, rape
CC      seed, sunflower, cotton, corn, tobacco, alfalfa, barley, oats,
CC      sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm,
CC      groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
CC      rice.
XX      Sequence 3957 BP; 1263 A; 790 C; 609 G; 1295 T; 0 other;
SQ
Query Match      100.0%; Score 1173; DB 22; Length 3957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATGAACCTTCCTTAAGCTCTTCCTTTCTACAGCTTTCTTTGTTGGTCAATACCTTCGTT 60
DB      1554 ATGAACCTTCCTTAAGCTCTTCCTTTCTACAGCTTTCTTTGTTGGTCAATACCTTCGTT 1613
QY      61 GCTGTTACTCAAGCTGCTGAGATCACCGCATTCCTCTCTACAAAGTAAAGTCTCTCGGT 120
DB      1614 GCTGTTACTCAAGCTGCTGAGATCACCGCATTCCTCTCTACAAAGTAAAGTCTCTCGGT 1673
QY      121 AAGGCGCTGAAGGAACATGACTTCTAGAAAGCTCTTGCAAGAAACAGATATGGCATC 180
DB      1674 AAGGCGCTGAAGGAACATGACTTCTAGAAAGCTCTTGCAAGAAACAGATATGGCATC 1733
QY      181 AGCAGCAAGTACTCCGGCTCGGTGAAGTGTGCTAGCGCACTTACCAACTACCTTGAT 240
DB      1734 AGCAGCAAGTACTCCGGCTCGGTGAAGTGTGCTAGCGCACTTACCAACTACCTTGAT 1793
QY      241 AGTCAATACCTTTGGGAAGATCTACTCTGGAAACCCCGCTCAAGAGTTCAACCGTTCTCTTT 300
DB      1794 AGTCAATACCTTTGGGAAGATCTACTCTGGAAACCCCGCTCAAGAGTTCAACCGTTCTCTTT 1853
QY      301 GATACGTGTTCTCTGACCTTCTGGGTTCCCTCTATCTCTGCAAGAAAGATGCGTGCAG 360
DB      1854 GATACGTGTTCTCTGACCTTCTGGGTTCCCTCTATCTCTGCAAGAAAGATGCGTGCAG 1913
QY      361 AACCAACCAAGATTGATCCGAGAAAGTCTTCAACCTTCCAGAACTTAGGCAAAACCTTG 420
DB      1914 AACCAACCAAGATTGATCCGAGAAAGTCTTCAACCTTCCAGAACTTAGGCAAAACCTTG 1973
QY      421 TCTATACCTAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 480
DB      1974 TCTATACCTAGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2033

```

```

QY      481 TCACAATTGTGACATTCAACAGACAGTACAGCTTACACCCAGAAACAGGTGATGTC 540
DB      2034 TCACAATTGTGACATTCAACAGACAGTACAGCTTACACCCAGAAACAGGTGATGTC 2093
QY      541 TTCACCTATGCAAGAAATTTGATGTCATCTTGTATGATGATACCCATTCGCTGCGTCAAG 600
DB      2094 TTCACCTATGCAAGAAATTTGATGTCATCTTGTATGATGATACCCATTCGCTGCGTCAAG 2153
QY      601 TACTGATACCTGTTGTTGACAAACATGATGAACCGACACTAGTACCTCAAGACTTGTTTC 660
DB      2154 TACTGATACCTGTTGTTGACAAACATGATGAACCGACACTAGTACCTCAAGACTTGTTTC 2213
QY      661 TCGGTTTACATGACAGGAATGCGCAGAGAGACATGCTCAAGCTTGAAGTATGATCA 720
DB      2214 TCGGTTTACATGACAGGAATGCGCAGAGAGACATGCTCAAGCTTGAAGTATGATCA 2273
QY      721 TCTTACTACAGAGATCTCTTCACTGCGGTTCCAGTCACTGTCAGCACTGCGCAATTC 780
DB      2274 TCTTACTACAGAGATCTCTTCACTGCGGTTCCAGTCACTGTCAGCACTGCGCAATTC 2333
QY      781 ACTGTGACAGTGTACCATGACAGCGGTGTTGTTGATGATGAAGTGAATGCAAGCT 840
DB      2334 ACTGTGACAGTGTACCATGACAGCGGTGTTGTTGATGATGAAGTGAATGCAAGCT 2393
QY      841 ATCTTGATACCGGTACGTTCCAGAGCTGTCGACCTTACAGCACTTCAACATTCAG 900
DB      2394 ATCTTGATACCGGTACGTTCCAGAGCTGTCGACCTTACAGCACTTCAACATTCAG 2453
QY      901 CAAGCTATTGAGCCACACAGAACCATGAGTGAAGTGAATGATGTCGACCAACCTT 960
DB      2454 CAAGCTATTGAGCCACACAGAACCATGAGTGAAGTGAATGATGTCGACCAACCTT 2513
QY      961 AGCTACATGCTTACAGTGTCTTTGATGATCAACGGCAAGATGATACCCACTGACCCCTTC 1020
DB      2514 AGCTACATGCTTACAGTGTCTTTGATGATCAACGGCAAGATGATACCCACTGACCCCTTC 2573
QY      1021 GCTATACACAGCCAGATCAAGGGTTCTGCACAGTGTATCCAGAGTGAAGAACCTATCC 1080
DB      2574 GCTATACACAGCCAGATCAAGGGTTCTGCACAGTGTATCCAGAGTGAAGAACCTATCC 2633
QY      1081 CAGAAATGATCTTGGAGATGTGTTCATTCTGATGATCTACAGCGTCTTGAACAGGACC 1140
DB      2634 CAGAAATGATCTTGGAGATGTGTTCATTCTGATGATCTACAGCGTCTTGAACAGGACC 2693
QY      1141 AACCAACCTCGTTGGGCTAGCTAAAGCAATCTGA 1173
DB      2694 AACCAACCTCGTTGGGCTAGCTAAAGCAATCTGA 2726

```

RESULT 3
AA030063
ID AA030063 standard; cDNA; 1098 BP.
XX
AC AA030063;
XX
DE 14-JUN-1992 (first entry)
XX
KW Sequence of proteinin cDNA in PCR 10001.
XX
KW Remnin; renin; enzyme; protease; ss.
XX
OS Bos taurus.
XX
FH Key
FT CDS 1..1098
XX Location/Qualifiers
XX PN EP73029-A.
XX PD 02-MAR-1993.
XX PF 19-AUG-1982; 82BP-0107601.

```

XX 24-AUG-1981; 81JP-0131631.
XX (BEPF/) BEPPU T.
XX Bepu T, Uozumi T, Nishimori K,
XX WPI: 1983-22976K/10.
XX P-PSDB; AAP30603.
XX plasmid contg. calf pro:rennin DNA - and transformed
XX microorganisms
XX Example; Page 20-23; 32pp; English.
XX The inventors claim recombinant plasmids contg. the cDNA of calf
XX prorennin. Specified plasmids are PCR 10001 and PCR2001 (contg. the
XX whole sequence plus the lac promoter region). Also new are
XX microorganisms transformed with the plasmids, esp. E. coli CR1 (ATCC
XX 391710) contg. plasmid PCR2001.
XX Sequence 1098 BP; 252 A; 326 C; 301 G; 219 T; 0 other;
XX
Query Match 4.2%; Score 49; DB 4; Length 1098;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 704
DB 581 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 629

RESULT 4
AAQ04683
ID AAQ04683 standard; DNA; 1098 BP.
XX
AC AAQ04683;
XX
XX 25-MAR-2003 (updated)
DT 05-OCT-1990 (first entry)
XX
XX Sequence encoding calf pro-rennin.
DE
XX Pro-rennin; ds.
XX
XX Bos taurus.
XX
XX JP02109984-A.
XX
XX 23-APR-1990.
PD
XX
XX 28-OCT-1981; 81JP-0302176.
PF
XX
XX 28-OCT-1981; 81JP-0302176.
PR
XX
XX (BEPF/) BEPPU T.
PA
XX
XX WPI: 1990-168358/22.
DR
XX P-PSDB; AAR05080.
XX
XX Complex plasmid and microbe - contains calf pro-rennin cDNA.
PT
XX
XX Disclosure; 32; 13pp; Japanese.
PS
XX
XX Protein product may be expressed in E.coli expression system from
CC plasmid pBR322.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct PR field.)
CC
XX Sequence 1098 BP; 252 A; 327 C; 300 G; 219 T; 0 other;
SQ
Query Match 4.2%; Score 49; DB 11; Length 1098;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;

```

```

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 704
DB 581 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 629

RESULT 5
AAN40295
ID AAN40295 standard; mRNA; 1175 BP.
XX
AC AAN40295;
XX
XX 04-FEB-1992 (first entry)
DT
XX
XX Sequence encoding a polypeptide displaying milk clotting activity.
DE
XX Cheese-making; recombinant protein; rennet substitute; milk clot; ss.
XX
XX Bos taurus.
XX
XX Key Location/Qualifiers
FH sig_peptide 21..69
FT /*tag= a
FT mat_peptide 70..1166
FT /*tag= b

XX EP123928-A.
XX
XX 07-NOV-1984.
XX
XX 30-MAR-1984; 84BP-0103551.
XX
XX 31-MAR-1983; 83US-0480860.
XX
XX (CODO-) CODON GENETIC ENG.
XX
XX Cashion LM, McCaman MT, Rice CW, Sias SR;
XX
XX WPI: 1984-277277/45.
DR
XX P-PSDB; AAP40559.
XX
XX Recombinant DNA coding for milk clotting polypeptide - which is
PT expressed in transformed bacteria
XX
XX Claim 6; Fig 2; 39pp; English.
XX
XX Also claimed is E.coli JM83/PLC7 (ATCC 39325) which is transformed
CC with PLC7 contg. the prorennin derived sequence fused in phase with
CC B-galactosidase. The PLC7 prorennin expression plasmid includes
CC sequences which code for both the pseudorennin and mature rennin
CC cleavage sites between AAs 28-29 and AAs 42-43, respectively.
XX
XX Sequence 1175 BP; 263 A; 355 C; 319 G; 238 U; 0 other;
SQ
Query Match 4.2%; Score 49; DB 5; Length 1175;
Best Local Similarity 75.5%; Pred. No. 1.5e-14;
Matches 37; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 704
DB 649 UGUUCUGGUUACAUGACAGGAUUGCCAGAGAGCAUGCUCACGCU 697

RESULT 6
AAQ20949
ID AAQ20949 standard; cDNA; 1175 BP.
XX
AC AAQ20949;
XX
XX 13-MAY-1992 (first entry)
DT
XX Prochymosin (prorennin) gene from calf.
DE

```

```

XX  Renet; zymogen; autocatalytic activation; pseudorennin; milk;
KM  clotting activity; ss.
XX
XX  Bos taurus.
OS
XX  Key      Location/Qualifiers
FH  CDS      21..1158
FT          /*tag= a
FT          /product= prorennin
FT          /note= "also known as prochymosin"
XX
XX  US5082775-A.
XX  PN
XX  PD  21-JAN-1992.
XX
XX  28-OCT-1988; 88US-0263927.
XX
XX  28-OCT-1988; 88US-0263927.
XX
XX  28-OCT-1988; 88US-0263927.
XX
XX  11-MAY-1984; 84US-0609495.
XX  12-DEC-1986; 86US-0940199.
XX  31-MAR-1983; 83US-0480860.
XX  28-APR-1986; 86US-0856700.
XX
XX  (BERL-) BERLEX LABS INC.
XX
XX  McCaman WT, King JF;
XX
XX  WPI; 1992-049149/06.
XX  DR  P-PSDB; AAR20730.
XX
XX  Isolating heterologous polypeptide from bacterial inclusion
XX  bodies - by lysing cells, extr. with nonionic detergent and sepg.
XX  insoluble polypeptide
XX
XX  Disclosure; Fig 6; 21pp; English.
XX
XX  The prorennin (prochymosin) sequence was obt'd from mRNA from
XX  unweaned calf's stomach (abomasum). To identify rennin-specific
XX  mRNA species, hybridisation with rennin specific probes was performed
XX  using probes derived from the published amino acid sequence of
XX  prorennin (Foltmann et al., J. Biol. Chem. 254, 8447-8456 (1979))
XX  (see AA023291.2). Reverse transcriptase was used to transcribe mRNA
XX  into a cDNA copy which was inserted into plasmid pBR322 and cloned
XX  into E. coli strain K-12 KM 294 (ATCC 31446). The prorennin
XX  specific clones were used to screen for recombinant plasmids.
XX  Only two clones 5G3 and 15G5 were found to contain the whole prorennin
XX  sequence, with clone 5G5 having its complete coding sequence (shown
XX  here). The sequence was the same as the published sequence except
XX  for two silent mutations in codons 274 and 336 respectively, and a
XX  single mutation in codon 302 which converts the sequence to the B form
XX  of rennin. The product of prorennin, rennin is an active component of
XX  rennet which is used to clot milk in the process of making cheese.
XX  See also AAQ20950.
XX
XX  Sequence 1175 BP; 261 A; 355 C; 319 G; 240 U; 0 other;
SQ
XX
XX  Query Match      4.2%; Score 49; DB 13; Length 1175;
XX  Best Local Similarity 75.5%; Pred. No. 1.5e-14;
XX  Matches 37; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
OY  656 TGTTCGCGTTTACATGACGAGATGCGCAGAGACATCTCAGCGT 704
XX  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB  649 UGUUCUGGCUUACUGACAGAAUGGCCAGAGACGACUGCUCGCU 697
XX
XX  RESULT 7
XX  ID  AAQ14051 standard; DNA; 1210 BP.
XX  AC  AAQ14051;
XX  XX
XX  DT  25-MAR-2003 (updated)

```

```

DT  06-JAN-1992 (first entry)
XX
XX  Renin gene.
DE
XX
XX  Prorennin; alpha-S1-casein gene; insulin-like growth factor I; IGF-I;
KM  mammary gland; ss.
XX
XX  Synthetic.
OS
XX
XX  EP451823-A.
XX
XX  16-OCT-1991.
XX
XX  10-APR-1991; 91EP-0105702.
XX
XX  19-APR-1980; 90DE-4012526.
XX  11-APR-1990; 90DE-4011751.
XX
XX  (CONE ) CONSORTIUM ELEKTROCHEM IND.
XX
XX  Hartl P, Brem G;
XX
XX  WPI; 1991-304858/42.
XX
XX  Recombinant DNA constructs for expressing protein in milk -
XX  contg. specific mammary gland transcription control region and
XX  signal sequence, providing high yield and easy prod. recovery
XX
XX  Disclosure; Page 21; 41pp; German.
XX
XX  The rennin gene was isolated using the probe represented in AAQ14775.
XX  It was used as heterologous peptide/protein together with parts of the
XX  alpha-S1-casein gene in the prodn. of DNA constructs. The heterologous
XX  peptide or protein may also be human insulin-like growth factor I.
XX  The constructs provide high yields of the protein with simple recovery
XX  from the milk. Activation of the gene occurs only in the mammary gland.
XX  See also AAQ14050, AAQ14774-77.
XX  (Updated on 25-MAR-2003 to correct PA field.)
XX
XX  Sequence 1210 BP; 278 A; 356 C; 324 G; 252 T; 0 other;
SQ
XX
XX  Query Match      4.2%; Score 49; DB 12; Length 1210;
XX  Best Local Similarity 100.0%; Pred. No. 1.5e-14;
XX  Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  656 TGTTCGCGTTTACATGACGAGATGCGCAGAGACATCTCAGCGT 704
XX  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB  643 TGTTCGCGTTTACATGACGAGATGCGCAGAGACATCTCAGCGT 691
XX
XX  RESULT 8
XX  ID  AAT03006 standard; DNA; 1240 BP.
XX  AC  AAT03006;
XX
XX  13-JUN-1996 (first entry)
DT
XX
XX  Chymosin open reading frame.
DE
XX
XX  a1.3; albino mutant; light-regulated; Neurospora; bread mould;
KM  heterologous gene; expression; control; chymosin; ss.
XX
XX  Mammalian sp.
XX
XX  WO9530739-A1.
XX
XX  16-NOV-1995.
XX
XX  09-MAY-1995; 95WO-US05716.
XX
XX  10-MAY-1994; 94US-0240372.
XX
XX

```

PA (UYHA-) UNIV HAWAII.
 XX Kato EK, Stuart WD;
 PI MPI, 1995-404108/51.
 DR Nucleic acid for expression of heterologous protein - contains
 XX albedo promoter for light induced expression in filamentous fungi
 PT Example 2, Fig 7, 26pp, English.
 PS
 XX The DNA is that of a mammalian gene (open reading frame) encoding
 CC chymosin. The gene was placed in operable linkage with the al-3
 CC promoter (see AAT03005). The al-3 gene controls the production of
 CC geranyl geranyl pyrophosphatase (GGPP) synthetase. GGPP is a precursor
 CC for carotenoids and xanthophylls. It has been shown that exposure to
 CC light increases the transcription level of GGPP synthetase 15-45 fold.
 CC Light activates a number of genes in the common bread mould, Neurospora.
 CC This can be used to regulate the expression of genes encoding
 CC heterologous proteins, e.g. chymosin, in recombinant production systems.
 CC Use of a light-regulated promoter is a simple and effective way to
 CC control expression and allows timing to be adapted to the physiological
 CC status of the host.
 CC
 SQ Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;
 Query Match 4.2%; Score 49; DB 16; Length 1240;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 656 TGTTCGCGTTTACATGACAGGAATGCGCAGAGAGCATGCTCAGCGT 704
 DB 700 TGTTCGCGTTTACATGACAGGAATGCGCAGAGAGCATGCTCAGCGT 748
 RESULT 9
 ID AAN40055 standard; DNA; 1278 BP.
 AC AAN40055;
 DT 02-FEB-1992 (first entry)
 XX Sequence of prochymosin gene.
 DE Prochymosin expression vector; E.coli trp operon; chymosin; ss.
 XX Bos taurus.
 OS
 XX Key Location/Qualifiers
 FH CDS 3..1130
 FT /*tag= a
 FT polyA_signal 1245..1250
 FT /*tag= b
 PN EP121775-A.
 XX 17-OCT-1984.
 PD
 XX 07-MAR-1984; 84BP-0102451.
 PF
 XX 09-MAR-1983; 83JP-0038439.
 DR
 XX (BEPP/) BEPPU T.
 PA
 XX BEPPU T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;
 PI Hataka M;
 XX WPI; 1984-258001/42.
 DR P-PSDB; AAP40078.
 XX
 PT Expression plasmid comprising prochymosin gene and vector -
 useful for transforming Escherichia coli for prochymosin prodn.

XX Disclosure; Fig 1: 59pp; English.
 XX
 CC The inventors claim the prochymosin gene comprising a nucleotide
 CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or
 CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant
 CC plasmids harboured by Escherichia coli strains deposited as FERM BP-
 CC 262, -263 and -264. Any portion of the nucleotide sequence as
 CC described in AAN40055 can be used. Also claimed is a vector derived
 CC from plasmid pBR322. Typically plasmid pBR501 is used. from pOCT 2.
 CC The transcriptional direction of pOCT 3 is opposite to that of
 CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter
 CC clockwise in pOCT 3.
 CC
 SQ Sequence 1278 BP; 309 A; 384 C; 326 G; 259 T; 0 other;
 Query Match 4.2%; Score 49; DB 5; Length 1278;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 656 TGTTCGCGTTTACATGACAGGAATGCGCAGAGAGCATGCTCAGCGT 704
 DB 613 TGTTCGCGTTTACATGACAGGAATGCGCAGAGAGCATGCTCAGCGT 661
 RESULT 10
 ID AAN30022 standard; DNA; 1289 BP.
 AC AAN30022;
 DT 25-MAR-2003 (updated)
 DT 25-APR-1992 (first entry)
 XX Sequence of veal chymosin gene.
 DE
 XX Protolytic enzyme; zymogen; rennin; chymosin; cheese making;
 KM microbial vector; ss.
 XX Bos taurus.
 OS
 XX Key Location/Qualifiers
 FH CDS 1..1140
 FT /*tag= a
 FT
 PN BE897201-A.
 XX 03-NOV-1983.
 PD
 XX 30-JUN-1983; 83BE-0017731.
 PF
 XX 13-APR-1983; 83US-0484539.
 PR 01-JUL-1982; 82US-0394433.
 XX (GENEX) GENEX CORP.
 PA
 XX WPI; 1983-820813/47.
 DR P-PSDB; AAP30013.
 XX
 PT Isolated chymosin or rennin and prochymosin genes - plasmid(s)
 PT which replicate in prokaryotic organisms, esp. Escherichia coli,
 FT and organisms used for chymosin biosynthesis
 XX
 PS Claim 7; Page 33-36; 43pp; French.
 CC The inventors claim isolated chymosin (rennin) and prochymosin genes
 CC from calves, and plasmids contg. the genes which are capable of
 CC replicating in a prokaryotic organism. The prokaryotic organism is
 CC pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).
 CC The microorganisms transformed by the plasmid are also claimed.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;

Query Match 4.2%; Score 49; DB 4; Length 1289;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 656 TGTTCGCGTTTACATGACAGAGATGGCCAGAGAGCATGCTCAGCT 704
 DB 623 TGTTCGCGTTTACATGACAGAGATGGCCAGAGAGCATGCTCAGCT 671

RESULT 11

AA91157
 ID AAN91157 standard; DNA; 1291 BP.

XX AAN91157;

DT 10-MAR-2003 (updated)
 DT 07-JUN-1990 (first entry)

XX Cloned sequence of (pro)chymosin.

XX Lactic acid bacteria; cheese; Streptococcus cremoris SK112; chymosin;
 KW proteinase; PSK112; ss.

XX Unidentified.

XX Key Location/Qualifiers
 FT precursor_RNA 29.1123
 FT /*tag= a
 FT /product=:Prochymosin

XX NL8701378-A.

XX PD 02-JAN-1989.

XX PF 12-JUN-1987; 87NL-0001378.

XX PR 12-JUN-1987; 87NL-0001378.

XX PA (NEZU-) NEDERL INS ZUIVELON.

XX PI Simons AFM; De Vos WM;

XX DR MPI; 1989-030097/04.

XX DR P-PSDB; AAP94144.

XX DNA fragment having region specific for lactic acid bacteria -
 PT is contained in plasmid in microorganism used in prodn. of
 PT protein and food prodn. eg cheese.

XX PS Disclosure; fig 2; 43pp; Dutch.

XX The DNA encoding prochymosin can be cloned into a plasmid (esp. from
 CC S. cremoris SK112) and used to produce large amts of the protein by
 CC recombinant DNA techniques. This could overcome the shortage of
 CC prochymosin due to shortage of calf stomachs and increasing cheese
 CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and
 CC buttermilk. See also AAN91158-N91160.
 CC (Updated on 10-MAR-2003 to add missing OS field.)

XX SQ Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

Query Match 4.2%; Score 49; DB 10; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 656 TGTTCGCGTTTACATGACAGAGATGGCCAGAGAGCATGCTCAGCT 704
 DB 609 TGTTCGCGTTTACATGACAGAGATGGCCAGAGAGCATGCTCAGCT 657

RESULT 12

AA91049

ID AAN30049 standard; cDNA; 1314 BP.

XX AAN30049;

XX 14-JUN-1992 (first entry)

DE cDNA sequence corresponding to one of the allelic forms (B) of
 DE bovine preprochymosin.

XX Chymosin; enzyme; rennet; cheese; ss.

XX Bos taurus.

XX Key Location/Qualifiers
 FH sig_peptide 24..40
 FT /*tag= a
 FT mat_peptide 41..1169
 FT /*tag= b

XX EP77109-A.

XX PD 20-APR-1983.

XX PF 13-OCT-1982; 82EP-0201272.

XX PR 14-OCT-1981; 81GB-0031004.

XX PA (UNIL) UNILEVER NV.

XX PI Maat J, Verrips CT, Ledebor AM, Edens L;

XX DR MPI; 1983-39656K/17.

XX DR P-PSDB; AAP30086.

XX DNA molecules comprising genes for preprochymosin - used to
 PT transform microorganisms to give strain producing the
 PT prepro-enzyme and its allelic and maturation forms

XX PS Claim 2; Fig 1; 53pp; English.

XX Preprochymosin is an intermediate (via prochymosin and
 CC pseudochymosin) for the enzyme chymosin, which is the essential milk-
 CC clotting component of rennet and is used in cheese manufacture.
 CC AAN30049 corresp. to mRNA isolated from the fourth stomach of a
 CC preterminant calf (abomasum, Frisian cow).

XX SQ Sequence 1314 BP; 309 A; 398 C; 338 G; 269 T; 0 other;

Query Match 4.2%; Score 49; DB 4; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 656 TGTTCGCGTTTACATGACAGAGATGGCCAGAGAGCATGCTCAGCT 704
 DB 652 TGTTCGCGTTTACATGACAGAGATGGCCAGAGAGCATGCTCAGCT 700

RESULT 13

AA920043
 ID AAN20043 standard; DNA; 1460 BP.

XX AAN20043;

XX DT 16-DEC-1992 (first entry)

XX DE Pre-prorennin-A gene DNA sequence.

XX Pre-pro-remnin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
 KW protease; milk-clotting enzyme; ss.

XX Bos taurus.

XX Key Location/Qualifiers

```
FT CDS 205..1350
FT /*tag= a
FT /label=
FT /note= "pre-prorennin-A gene"
XX
XX GB2091271-A.
XX
XX 28-JUL-1982.
XX
XX 15-JAN-1982; 82GB-0001120.
XX
XX 01-DEC-1981; 81US-0325481.
XX 16-JAN-1981; 81US-0225717.
XX
XX (COLB ) COLLABORATIVE RES INC.
XX
XX Alford BL, Mao J, Moir DT;
XX
XX WPI; 1982-62028E/30 (62028B).
XX P-PSDB; AAP20038.
XX
XX Transformed cells producing rennin and its precursors - contg.
XX appropriate recombinant DNA material
XX
XX Disclosure; Table 1; 39pp; English.
XX
XX Bases 1-204 and 1351 to 1460 are attached to the pre-prorennin but
XX can be removed and are not essential to use of the gene in
XX expression. The gene may be ligated into plasmid pCGE21 and
XX expressed in E. coli. The resulting expressed enzyme is a well
XX known milk-clotting enzyme used in cheese-making.
XX
XX Sequence 1460 BP; 329 A; 440 C; 397 G; 294 T; 0 other;
SQ
Query Match 4.2%; Score 49; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 656 TGTTCGCGTTTACATGACAGGAATGCCAGAGAGACATGCTCAGCT 704
DB 833 TGTTCGCGTTTACATGACAGGAATGCCAGAGAGACATGCTCAGCT 881
RESULT 14
AAN40180
ID AAN40180 standard; CDNA; 1460 BP.
XX
XX AAN40180;
XX
XX 25-MAR-2003 (updated)
XX 25-JAN-1992 (first entry)
XX
XX Sequence of recombinant CGF4 carrying the rennin coding sequence.
XX
XX Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae;
XX es.
XX
XX Bos taurus.
XX
XX Key Location/Qualifiers
XX CDS 205..1350
XX FT /**tag= a
XX
XX GB2137208-A.
XX
XX 03-OCT-1984.
XX
XX 28-FEB-1984; 84GB-0005129.
XX
XX 28-FEB-1983; 83US-0470911.
XX
XX (COLB ) COLLABORATIVE RES INC.
XX
```

```
PI Botstein D, Davis RW, Fink GR, Tauntonrig A, Knowlton RG, Mao JI;
PI Moir DT, Goff CG;
XX
XX WPI; 1984-245517/40.
XX P-PSDB; AAP40218.
XX
XX DNA segment contg. GAL1 promoter linked to gene - useful for
XX direction of expression of the gene in yeast cell
XX
XX Example; Table 4, Page 21-23; 35pp; English.
XX
XX The inventors claim a DNA segment contg. GAL1 promoter linked to
XX gene - useful for direction of expression of the gene in yeast
XX cell. The recombinant material carrying a GAL1 promoter of the yeast
XX galactokinase gene may be used in expressing a desired protein, esp.
XX bovine growth hormone, interferon, prorennin or preprorennin, in the
XX yeast cell. Strains of Saccharomyces cerevisiae producing the
XX polypeptides are produced. Yeast strains deposited as ATCC 20643,
XX 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
XX 528, resp. are new.
XX (Updated on 25-MAR-2003 to correct 'P' field.)
XX (Updated on 25-MAR-2003 to correct 'I' field.)
XX
XX Sequence 1460 BP; 328 A; 440 C; 399 G; 293 T; 0 other;
SQ
Query Match 4.2%; Score 49; DB 5; Length 1460;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 656 TGTTCGCGTTTACATGACAGGAATGCCAGAGAGACATGCTCAGCT 704
DB 833 TGTTCGCGTTTACATGACAGGAATGCCAGAGAGACATGCTCAGCT 881
RESULT 15
AAN91188
ID AAN91188 standard; DNA; 2727 BP.
XX
XX AAN91188;
XX
XX 25-MAR-2003 (updated)
XX 15-JUN-1990 (first entry)
XX
XX BamHI/SalI insert of Kluyveromyces plasmid pAB309.
XX
XX Kluyveromyces; pDM100PC; chymosin; tissue plasminogen activator.
XX
XX Kluyveromyces lactis.
XX
XX Key Location/Qualifiers
XX CDS 409..1781
XX FT /**tag= a
XX
XX EP301670-A.
XX
XX 01-FEB-1989.
XX
XX 28-JUL-1988; 88EP-0201632.
XX
XX 28-JUL-1987; 87US-0078539.
XX
XX (KONN ) GIST-BROCADES NV.
XX
XX Vandenberg JA, Vanooyen AJJ, Rietveld K;
XX
XX WPI; 1989-033565/05.
XX P-PSDB; AAP94376.
XX
XX Kluyveromyces host cells for producing polypeptide(s) -
XX used for highly efficient prodn. of eg chymosin tissue
XX plasminogen activator or human serum albumin.
XX
XX Disclosure; Page 7; 56pp; English.
XX
```

XX BamHI/SalI insert, incorporated into plasmids pAB309 with
 CC promoters, terminators and a G418 resistance marker fused to a ADH1
 CC promoter from *S.cerevisiae*.
 CC The Kluyveromyces expression systems provide highly efficient secretion
 CC and processing of a wide variety of proteins.
 CC Sequences identical to those published in EP01669.
 CC (Updated on 25-MAR-2003 to correct PD field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)
 XX

SQ Sequence 2727 BP; 752 A; 654 C; 572 G; 749 T; 0 other;

Query Match 4.2%; Score 49; DB 10; Length 2727;
 Best Local Similarity 100.0%; Pred.No. 1.5e-14;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTTCGCGTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 704
 |||||
 Db 1268 TGTTCGCGTTACATGACAGGAATGGCCAGAGAGCATGCTCAGCGT 1316

Search completed: February 5, 2004, 06:07:28
 Job time : 396 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 5, 2004, 05:20:22 ; Search time 92 Seconds

(without alignments)
5627.635 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173

Sequence: 1 atgactctcttaagctctt.....ggctagctaaagcaatctga 1173

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4.2	1240	1	US-08-240-372-2	Sequence 2, Appl
2	4.2	2732	2	US-08-846-021A-6	Sequence 6, Appl
3	3.2	2732	6	5217891-14	Patent No. 5217891
4	2.0	828	6	5200327-6	Patent No. 5200327
5	2.0	1.7	832	1	US-08-318-193-15
6	2.0	1.7	838	1	US-08-318-193-13
7	2.0	1.7	846	6	5200327-5
8	1.6	689	4	US-09-149-476-263	Sequence 263, App
9	1.6	1751	4	US-09-149-476-110	Sequence 110, App
10	1.6	13953	4	US-09-138-884-3	Sequence 3, Appl
11	1.5	1746	4	US-09-107-532A-1731	Sequence 1731, Ap
12	1.4	420	1	US-08-318-970B-38	Sequence 38, Appl
13	1.4	420	1	US-08-318-970B-47	Sequence 47, Appl
14	1.4	550	4	US-09-359-301A-24	Sequence 24, Appl
15	1.4	555	4	US-09-071-035-13	Sequence 13, Appl
16	1.4	738	2	US-08-738-462-1	Sequence 1, Appl
17	1.4	738	5	PCT-US94-07587-1	Sequence 1, Appl
18	1.4	2576	1	US-08-471-033-35	Sequence 35, Appl
19	1.4	2576	2	US-08-471-044-35	Sequence 35, Appl
20	1.4	2576	2	US-08-463-483A-35	Sequence 35, Appl
21	1.4	2576	2	US-08-471-046A-35	Sequence 35, Appl
22	1.4	2576	2	US-08-470-566B-35	Sequence 35, Appl
23	1.4	2576	2	US-08-469-334-35	Sequence 35, Appl
24	1.4	2576	2	US-09-300-529-35	Sequence 35, Appl
25	1.4	2576	4	US-09-252-991A-8530	Sequence 35, Appl
26	1.4	2655	1	US-08-471-033-17	Sequence 17, Appl
27	1.4	2655	1	US-08-471-033-26	Sequence 26, Appl

28	1.4	2655	2	US-08-471-044-17	Sequence 17, Appl
29	1.4	2655	2	US-08-471-044-26	Sequence 26, Appl
30	1.4	2655	2	US-08-463-483A-17	Sequence 17, Appl
31	1.4	2655	2	US-08-463-483A-26	Sequence 26, Appl
32	1.4	2655	2	US-08-471-046A-17	Sequence 17, Appl
33	1.4	2655	2	US-08-471-046A-26	Sequence 26, Appl
34	1.4	2655	2	US-08-470-566B-17	Sequence 17, Appl
35	1.4	2655	2	US-08-470-566B-26	Sequence 26, Appl
36	1.4	2655	2	US-08-469-334-17	Sequence 17, Appl
37	1.4	2655	2	US-08-469-334-26	Sequence 26, Appl
38	1.4	2655	3	US-09-300-529-17	Sequence 17, Appl
39	1.4	2655	3	US-09-300-529-26	Sequence 26, Appl
40	1.4	2754	4	US-09-252-991A-8530	Sequence 8530, Ap
41	1.4	3471	2	US-08-715-568A-2	Sequence 2, Appl
42	1.4	4031	1	US-08-471-033-49	Sequence 49, Appl
43	1.4	4031	2	US-08-471-044-49	Sequence 49, Appl
44	1.4	4031	2	US-08-463-483A-49	Sequence 49, Appl
45	1.4	4031	2	US-08-471-046A-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1
US-08-240-372-2
; Sequence 2, Application US/08240372
; Patent No. 5741665
; GENERAL INFORMATION:
; APPLICANT: KATO, ELIE K.
; TITLE OF INVENTION: STUART, W. DORSEY
; TITLE OF INVENTION: LIGHT-REGULATED PROMOTERS FOR PRODUCTION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESS: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,372
; FILING DATE: 10-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REFERENCE/DOCKET NUMBER: 29,959
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELE: 90-4030
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-240-372-2

Query Match 4.2%; Score 49; DB 1; Length 1240;
Best Local Similarity 100.0%; Pred. No. 9.4e-16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 TGTCTCGCTTATCATGACGAGATGCGCCAGAGCATCTACAGCT 704
DB 700 TGTCTCGCTTATCATGACGAGATGCGCCAGAGCATCTACAGCT 748

RESULT 2
US-08-846-021A-6
; Sequence 6, Application US/08846021A
; Patent No. 5948682
; GENERAL INFORMATION:
; APPLICANT: MOLONEY, Maurice M.
; TITLE OF INVENTION: Preparation of Heterologous Proteins on
; TITLE OF INVENTION: Oil Bodies
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERSKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846.021A
; FILING DATE: April 25, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 9369-039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 850..1206
; ;
; NAME/KEY: CDS
; LOCATION: 1444..2729
; ;
; US-08-846-021A-6
; ;
Query Match 4.2%; Score 49; DB 2; Length 2733;
Best Local Similarity 100.0%; Pred. No. 9.5e-16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; ;
QY 656 TGTTCGCGTTTACATGACAGATGCGCCAGAGACATGCTCAGCCT 704
Db 2210 TGTTCGCGTTTACATGACAGATGCGCCAGAGACATGCTCAGCCT 2258
; ;
RESULT 3
5217891-14
; Patent No. 5217891
; APPLICANT: BRAKE, ANTHONY J.; VAN DEN BERG, JOHAN A.
; TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
; A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
; POLYPEPTIDES
; NUMBER OF SEQUENCES: 23
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507.398
; FILING DATE: 09-APR-1990
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 78,551
; FILING DATE: 28-JUL-1987
; SEQ ID NO: 14:
; LENGTH: 2732
; 5217891-14

Query Match 3.2%; Score 38; DB 6; Length 2732;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; ;
QY 778 TTCACGTGACAGTGTACATCAGCGGTGTGTGT 815
Db 1391 TTCACGTGACAGTGTACATCAGCGGTGTGTGT 1428
; ;
RESULT 4
5200327-6
; Patent No. 5200327
; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
; STREPTOMYCES
; NUMBER OF SEQUENCES: 24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/224.568
; FILING DATE: 26-JUL-1988
; SEQ ID NO: 6:
; LENGTH: 828
; 5200327-6
; ;
Query Match 1.7%; Score 20; DB 6; Length 828;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; ;
QY 358 AAGAACCAACCAAGATTGCA 377
Db 795 AAGAACCAACCAAGATTGCA 814
; ;
RESULT 5
US-08-118-193-15
; Sequence 15, Application US/08118193
; Patent No. 564163
; GENERAL INFORMATION:
; APPLICANT: GARVIN, Robert T.
; APPLICANT: MALEK, Lawrence T.
; TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
; TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318.193
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.314
; FILING DATE:
; APPLICATION NUMBER: US 07/224.568
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/116 CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300

TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid:
DESCRIPTION: Synthetic DNA oligonucleotide
FEATURE:
NAME/KEY: CDS
LOCATION: 399..830
US-08-318-193-15

Query Match 1.7%; Score 20; DB 1; Length 832;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 AAGAACCACCAAGATTGCA 377
|||||
Db 795 AAGAACCACCAAGATTGCA 814

RESULT 6
US-08-318-193-13
Sequence 13, Application US/08318193
Patent No. 5641663
GENERAL INFORMATION:
APPLICANT: GARVIN, Robert T.
APPLICANT: MALEK, Lawrence T.
TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,193
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,314
FILING DATE:
APPLICATION NUMBER: US 07/224,568
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/116 CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 838 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid:
DESCRIPTION: Synthetic DNA oligonucleotide
FEATURE:

NAME/KEY: CDS
LOCATION: 399..836
US-08-318-193-13

Query Match 1.7%; Score 20; DB 1; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 AAGAACCACCAAGATTGCA 377
|||||
Db 801 AAGAACCACCAAGATTGCA 820

RESULT 7
5200327-5
Patent No. 5200327
APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
STREPTOMYCES
NUMBER OF SEQUENCES: 24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/224,568
FILING DATE: 26-JUL-1988
SEQ ID NO: 5;
LENGTH: 846
5200327-5

Query Match 1.7%; Score 20; DB 6; Length 846;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 AAGAACCACCAAGATTGCA 377
|||||
Db 801 AAGAACCACCAAGATTGCA 820

RESULT 8
US-09-149-476-263/c
Sequence 263, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633

[illegible]

EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048, 964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057, 650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056, 884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057, 669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049, 610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061, 060
EARLIER FILING DATE: 1997-10-02

Query Match 1.6%; Score 19; DB 4; Length 689;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1068 TGAGAACCATCCCGAGAA 1086
Db 323 TGAGAACCATCCCGAGAA 305

RESULT 9
US-09-149-476-110/c
Sequence 110, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002p1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040, 162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040, 333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038, 621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040, 626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040, 334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040, 336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040, 163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047, 600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 618
EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047, 503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047, 601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043, 580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043, 315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048, 974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056, 886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056, 637

EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 1.6%; Score 19; DB 4; Length 1751;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1068 TGAGAACCATTCGCCAGAA 1086
Db 1101 TGAGAACCATTCGCCAGAA 1083

RESULT 10
US-09-738-884-3/c
Sequence 3, Application US/09738884
Patent No. 6391606
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
FILE REFERENCE: CL000849
CURRENT APPLICATION NUMBER: US/09/738,884
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 13953
TYPE: DNA
ORGANISM: Human
US-09-738-884-3

Query Match 1.6%; Score 19; DB 4; Length 13953;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 GCATTCCTCTCTCAAGG 107
Db 10698 GCATTCCTCTCTCAAGG 10680

RESULT 11
US-09-107-532A-1731
Sequence 1731, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
City: Waltham
State: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1731:
SEQUENCE CHARACTERISTICS:
LENGTH: 1746 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1..1746
SEQUENCE DESCRIPTION: SEQ ID NO: 1731:
US-09-107-532A-1731

Query Match 1.5%; Score 18; DB 4; Length 1746;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 AGGCGCTGAGGACATG 139
Db 509 AGGCGCTGAGGACATG 526

RESULT 12
US-08-318-970B-38
Sequence 38, Application US/08318970B
Patent No. 3589573
GENERAL INFORMATION:
APPLICANT: Hideaki HAGIWARA, et al.
TITLE OF INVENTION: AMINO ACID SEQUENCES OF ANTI-IDIOTYPIC
TITLE OF INVENTION: ANTIBODIES AGAINST ANTI-CANCER HUMAN MONOCLONAL ANTIBODY
TITLE OF INVENTION: AND DNA BASE SEQUENCES ENCODING THOSE SEQUENCES
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: Dell System 210; Intel 80 285 Microprocessor
OPERATING SYSTEM: MS DOS 3.3; Version 5.1
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,970B
FILING DATE: October 6, 1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: S-2371
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: mouse
FEATURE:
NAME/KEY: Idio 27 L chain
US-08-318-970B-38

Query Match 1.4%; Score 17; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 794 TCACCATCAGCGGTGTG 810
Db 218 TCACCATCAGCGGTGTG 234

RESULT 13
US-08-318-970B-47
Sequence 47, Application US/08318970B
Patent No. 5589573
GENERAL INFORMATION:
APPLICANT: Hideaki HAGIWARA, et al.
TITLE OF INVENTION: AMINO ACID SEQUENCES OF ANTI-IDIOTYPIC
TITLE OF INVENTION: ANTIBODIES AGAINST ANTI-CANCER HUMAN MONOCLONAL ANTIBODY
TITLE OF INVENTION: AND DNA BASE SEQUENCES ENCODING THOSE SEQUENCES
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: Dell System 210; Intel 80 285 Microprocessor
OPERATING SYSTEM: MS DOS 3.3
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,970B
FILING DATE: October 6, 1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: S-2371
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 420
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: mouse
FEATURE:
NAME/KEY: Clone 27KA2
US-08-318-970B-47

Query Match 1.4%; Score 17; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 794 TCACCATCAGCGGTGTG 810
DB 308 TCACCATCAGCGGTGTG 324

RESULT 14
US-09-359-301A-24
Sequence 24, Application US/09359301A
Patent No. 6426185
GENERAL INFORMATION:
APPLICANT: Kumagai, Monto H.
APPLICANT: della-Cioppa, Guy R.
APPLICANT: Erwin, Robert L.
APPLICANT: McGee, David R.
TITLE OF INVENTION: METHOD OF DETERMINING THE PRESENCE OF A
TITLE OF INVENTION: TRAIT IN A PLANT BY TRANSFECTING A NUCLEIC ACID SEQUENCE OF
TITLE OF INVENTION: A DONOR PLANT INTO A DIFFERENT HOST PLANT IN AN ANTI-SENSE
FILE REFERENCE: 008010137US04
CURRENT APPLICATION NUMBER: US/09/359,301A
CURRENT FILING DATE: 1999-07-21
NUMBER OF SEQ ID NOS: 42
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 550
TYPE: DNA
ORGANISM: Oryza sativa
US-09-359-301A-24

Query Match 1.4%; Score 17; DB 4; Length 550;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 ACCGCTGCTGAGATCACC 87
DB 406 ACCGCTGCTGAGATCACC 422

RESULT 15
US-09-071-035-13/c
Sequence 13, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis polynucleotides and polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-13

Query Match 1.4%; Score 17; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 AAGAGCAATGCTGCAA 359
DB 60 AAGAGCAATGCTGCAA 44

Search completed: February 5, 2004, 08:14:30
Job time: 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 06:01:29 ; Search time 489 Seconds
(without alignments)
8836.199 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173
Sequence: 1 atgaacttccttaagtcctt.....ggctacgtaacgaatcga 1173

Scoring table: OLIGO_NUC

Gapop_60.0 , Gapext 60.0

Searched: 2449703 seqs, 1841816367 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	9.0	1041	US-10-324-131-10	Sequence 10, Appl
2	49	4.2	2733	US-09-893-525-6	Sequence 6, Appl
3	34	2.9	1096	US-10-322-746-1	Sequence 18, Appl
4	33	2.8	1119	US-10-332-746-3	Sequence 3, Appl
5	23	2.0	609	US-09-770-149-886	Sequence 886, App
6	23	2.0	1521	US-09-938-842A-2245	Sequence 2245, App
7	23	2.0	1053	US-09-938-842A-2245	Sequence 2245, App
8	22	1.9	1080	US-10-322-746-16	Sequence 16, Appl
9	22	1.9	1101	US-10-322-746-14	Sequence 14, Appl
10	22	1.9	1134	US-10-332-746-12	Sequence 12, Appl
11	22	1.9	2756	US-09-887-576-449	Sequence 449, App
12	22	1.6	231	US-09-815-343-50	Sequence 50, Appl
13	19	1.6	503	US-10-027-632-106390	Sequence 106390,
14	19	1.6	503	US-10-027-632-106390	Sequence 106390,
15	19	1.6	503	US-10-027-632-106390	Sequence 106390,

16	19	1.6	556	10	US-09-998-598-1222	Sequence 1222, Ap
17	19	1.6	689	11	US-09-809-391-263	Sequence 263, App
18	19	1.6	689	11	US-09-882-171-263	Sequence 263, App
19	19	1.6	714	13	US-10-027-632-22346	Sequence 22346, A
20	19	1.6	714	14	US-10-027-632-22346	Sequence 22346, A
21	19	1.6	1751	11	US-09-809-391-110	Sequence 110, App
22	19	1.6	1751	13	US-09-882-171-110	Sequence 110, App
23	19	1.6	4170	11	US-09-919-039-221	Sequence 221, App
24	19	1.6	4170	15	US-10-168-425-23	Sequence 23, Appl
25	19	1.6	13440	15	US-10-213-948-12	Sequence 12, Appl
26	19	1.6	13953	14	US-10-086-961-3	Sequence 3, Appl
27	19	1.5	60	13	US-09-908-975-23544	Sequence 23544, A
28	18	1.5	527	11	US-09-918-995-31599	Sequence 31599, A
29	18	1.5	558	13	US-10-027-632-207782	Sequence 207782,
30	18	1.5	558	13	US-10-027-632-207782	Sequence 207784,
31	18	1.5	558	13	US-10-027-632-207785	Sequence 207785,
32	18	1.5	558	14	US-10-027-632-207782	Sequence 207782,
33	18	1.5	558	14	US-10-027-632-207784	Sequence 207784,
34	18	1.5	558	14	US-10-027-632-207785	Sequence 207785,
35	18	1.5	632	13	US-10-027-632-200742	Sequence 200742,
36	18	1.5	632	13	US-10-027-632-200743	Sequence 200743,
37	18	1.5	632	14	US-10-027-632-200742	Sequence 200743,
38	18	1.5	632	14	US-10-027-632-200743	Sequence 200743,
39	18	1.5	648	11	US-09-791-279-83	Sequence 83, Appl
40	18	1.5	1117	13	US-10-027-632-207783	Sequence 207783,
41	18	1.5	1117	14	US-10-027-632-207783	Sequence 207783,
42	18	1.5	1365	13	US-10-027-632-251427	Sequence 251427,
43	18	1.5	1365	13	US-10-027-632-251428	Sequence 251428,
44	18	1.5	1365	13	US-10-027-632-251429	Sequence 251429,
45	18	1.5	1365	13	US-10-027-632-251430	Sequence 251430,

ALIGNMENTS

RESULT 1
US-10-324-131-10
Sequence 10, Application US/10324131
Publication No. US2003017537A1
GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: Expression of Epidermal Growth Factor in Plant Seeds
FILE REFERENCE: 9369-251
CURRENT APPLICATION NUMBER: US/10/324,131
PRIOR FILING DATE: 2002-12-20
PRIOR APPLICATION NUMBER: US 07/659,835
PRIOR FILING DATE: 1991-02-22
PRIOR APPLICATION NUMBER: US 08/142,418
PRIOR FILING DATE: 1993-11-16
PRIOR APPLICATION NUMBER: US 08/366,783
PRIOR FILING DATE: 1994-12-30
PRIOR APPLICATION NUMBER: US 08/846,021
PRIOR FILING DATE: 1997-04-25
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 1041
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Oleosin-BGF
US-10-324-131-10
Query Match 9.0%; Score 105; DB 13; Length 1041;
Best Local Similarity 100.0%; Pred. No. 4.3e-47;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 76 GCTGAGATCACCAGCCGATTCCTCTTACCAAGGTAAGTCTTCGTAAGCGCTGAAGAA 135
DB 763 GCTGAGATCACCAGCCGATTCCTCTTACCAAGGTAAGTCTTCGTAAGCGCTGAAGAA 822
QY 136 CATGACTTCTAGAAGACTCTTCGAGAAACAAGTATGATC 180

Db 823 CATGACCTTTAGACAGCTTTCTTCGAGAAACAAACAGTATGGCATC 867

RESULT 2

US-09-893-525-6
; Sequence 6, Application US/09893525
; Publication No. US20030126631A1
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, Gijb
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210,843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; PRIOR FILING DATE: 1991-02-22
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (850)..(1203)
; OTHER INFORMATION:
; NAME/KEY: CDS
; LOCATION: (1444)..(2727)
; OTHER INFORMATION:
US-09-893-525-6

Query Match 4.2%; Score 49; DB 11; Length 2733;
Best Local Similarity 100.0%; Pred. No. 3.9e-16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 656 TGTTCCTCGTTTACATGACGAGATGCGCAGAGACATGCTCAGCGCT 704
Db 2210 TGTTCCTCGTTTACATGACGAGATGCGCAGAGACATGCTCAGCGCT 2258

RESULT 3

US-10-322-746-1
; Sequence 1, Application US/10322746
; Publication No. US20030166162A1
; GENERAL INFORMATION:
; APPLICANT: van Rooijen, Gijb
; APPLICANT: Alcantara, Joemel
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Method for Cleavage of Fusion Proteins
; FILE REFERENCE: 9369-227
; CURRENT APPLICATION NUMBER: US/10/322,746
; CURRENT FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1096
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GST-Pro-Hirudin fusion
US-10-322-746-1

Query Match 2.9%; Score 34; DB 13; Length 1096;

Best Local Similarity 100.0%; Pred. No. 7.6e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 CAGTATGCATCAGCAGCAAGTACTCCGGCTTCG 202
Db 802 CAGTATGCATCAGCAGCAAGTACTCCGGCTTCG 835

RESULT 4

US-10-322-746-3
; Sequence 3, Application US/10322746
; Publication No. US20030166162A1
; GENERAL INFORMATION:
; APPLICANT: van Rooijen, Gijb
; APPLICANT: Alcantara, Joemel
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Method for Cleavage of Fusion Proteins
; FILE REFERENCE: 9369-227
; CURRENT APPLICATION NUMBER: US/10/322,746
; CURRENT FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: His-Pro-cGH
US-10-322-746-3

Query Match 2.8%; Score 33; DB 13; Length 1119;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 CAGTATGCATCAGCAGCAAGTACTCCGGCTTCG 201
Db 220 CAGTATGCATCAGCAGCAAGTACTCCGGCTTCG 252

RESULT 5

US-09-770-149-886/C
; Sequence 886, Application US/09770149
; Patent No. US2002059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Olang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2024 (PAPA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 886

LENGTH: 609
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-149-886

Query Match 2.0%; Score 23; DB 9; Length 609;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGGAGATGTTCAT 1109
DB 389 TGGATCTTGGGAGATGTTCAT 367

RESULT 6
US-09-938-842A-2245
Sequence 2245, Application US/09938842A
Patent No. US2002016037B1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2245
LENGTH: 1521
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2245

Query Match 2.0%; Score 23; DB 10; Length 1521;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGGAGATGTTCAT 1109
DB 1435 TGGATCTTGGGAGATGTTCAT 1457

RESULT 7
US-09-938-842A-2245
Sequence 2245, Application US/09938842A
Publication No. US20040009476A9
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2245
LENGTH: 1521

TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2245

Query Match 2.0%; Score 23; DB 12; Length 1521;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGGAGATGTTCAT 1109
DB 1435 TGGATCTTGGGAGATGTTCAT 1457

RESULT 8
US-10-322-746-18
Sequence 18, Application US/10322746
Publication No. US20030166162A1
GENERAL INFORMATION:
APPLICANT: van Rooijen, Gijb
APPLICANT: Alcantara, Joanel
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: Method for Cleavage of Fusion Proteins
FILE REFERENCE: 9369-227
CURRENT APPLICATION NUMBER: US/10/322,746
CURRENT FILING DATE: 2002-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.1
SEQ ID NO 18
LENGTH: 1053
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: GST-KLIP14-cystatin
US-10-322-746-18

Query Match 1.9%; Score 22; DB 13; Length 1053;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 CAGTATGGCATCAGCAGCAACT 190
DB 721 CAGTATGGCATCAGCAGCAACT 742

RESULT 9
US-10-322-746-16
Sequence 16, Application US/10322746
Publication No. US20030166162A1
GENERAL INFORMATION:
APPLICANT: van Rooijen, Gijb
APPLICANT: Alcantara, Joanel
APPLICANT: Moloney, Maurice M.
TITLE OF INVENTION: Method for Cleavage of Fusion Proteins
FILE REFERENCE: 9369-227
CURRENT APPLICATION NUMBER: US/10/322,746
CURRENT FILING DATE: 2002-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 1080
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: GST-KLIP12-cystatin
US-10-322-746-16

Query Match 1.9%; Score 22; DB 13; Length 1080;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 CAGTATGGCATCAGCAGCAACT 190
DB 748 CAGTATGGCATCAGCAGCAACT 769

RESULT 10
US-10-322-746-14
; Sequence 14, Application US/10322746
; Publication No. US20030166162A1
; GENERAL INFORMATION:
; APPLICANT: van Rooijen, Gijb
; APPLICANT: Alcantara, Joemel
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Method for Cleavage of Fusion Proteins
; FILE REFERENCE: 9369-227
; CURRENT APPLICATION NUMBER: US/10/322,746
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 14
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GST-KLIP11-cystatin
US-10-322-746-14

Query Match 1.9%; Score 22; DB 13; Length 1101;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 CAGTATGGCATCAGCAGCAAGT 190
Db 769 CAGTATGGCATCAGCAGCAAGT 790

RESULT 11
US-10-322-746-12
; Sequence 12, Application US/10322746
; Publication No. US20030166162A1
; GENERAL INFORMATION:
; APPLICANT: van Rooijen, Gijb
; APPLICANT: Alcantara, Joemel
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Method for Cleavage of Fusion Proteins
; FILE REFERENCE: 9369-227
; CURRENT APPLICATION NUMBER: US/10/322,746
; CURRENT FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 12
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GST-KLIP4-cystatin
US-10-322-746-12

Query Match 1.9%; Score 22; DB 13; Length 1134;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 CAGTATGGCATCAGCAGCAAGT 190
Db 802 CAGTATGGCATCAGCAGCAAGT 823

RESULT 12
US-09-887-576-449
; Sequence 449, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.

; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 449
; LENGTH: 2756
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-449

Query Match 1.9%; Score 22; DB 10; Length 2756;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1088 GGATCTTGGGAGATGTGTTTCAT 1109
Db 2671 GGATCTTGGGAGATGTGTTTCAT 2692

RESULT 13
US-09-815-343-50/c
; Sequence 50, Application US/09815343
; Patent No. US2001005596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jianshun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 231
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(231)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-50

Query Match 1.6%; Score 19; DB 9; Length 231;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1072 AACCATTCGAGAAATGCA 1090
Db 153 AACCATTCGAGAAATGCA 135

RESULT 14
US-10-027-632-106390
; Sequence 106390, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129

Search completed: February 5, 2004, 08:25:27
Job time : 494 secs

CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: fastseq for windows Version 4.0
SEQ ID NO 106390
LENGTH: 503
TYPE: DNA
ORGANISM: Human
US-10-027-632-106390

Query Match 1.6%; Score 19; DB 13; Length 503;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GCATTCCTCTCTACAAAGG 107
|||||
Db 436 GCATTCCTCTCTACAAAGG 454

RESULT 15
US-10-027-632-106390
Sequence 106390, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: fastseq for windows Version 4.0
SEQ ID NO 106390
LENGTH: 503
TYPE: DNA
ORGANISM: Human
US-10-027-632-106390

Query Match 1.6%; Score 19; DB 14; Length 503;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GCATTCCTCTCTACAAAGG 107
|||||
Db 436 GCATTCCTCTCTACAAAGG 454

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 5, 2004, 05:19:09 ; Search time 2885 Seconds
(without alignments)
9881.855 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173
Sequence: 1 atgactctcctaagctctt.....ggtcagcctaagcaatctga 1173

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4.9	4.2	307	14	CB221243
2	4.9	4.2	383	12	BG938086
3	4.9	4.2	430	12	BG937697
4	4.9	4.2	472	12	BG938320

5	4.3	3.7	399	12	BG937723
6	4.2	3.6	410	14	CB220875
7	3.8	3.2	297	14	CB220520
8	3.8	3.2	443	14	CB220074
9	3.2	2.7	204	29	BZ911661
10	2.3	2.0	325	9	AV532492
11	2.3	2.0	339	14	T20903
12	2.3	2.0	380	9	AV538262
13	2.3	2.0	391	9	AA969042
14	2.3	2.0	396	9	AB801400
15	2.3	2.0	398	9	AB801901
16	2.3	2.0	398	9	AB802605
17	2.3	2.0	404	9	AA411567
18	2.3	2.0	405	9	AB806186
19	2.3	2.0	405	9	AB814825
20	2.3	2.0	408	9	AV518232
21	2.3	2.0	408	9	AV798981
22	2.3	2.0	409	9	AV806470
23	2.3	2.0	410	9	AV802625
24	2.3	2.0	410	9	AV803851
25	2.3	2.0	411	9	AV559941
26	2.3	2.0	413	9	AV802692
27	2.3	2.0	420	9	AV805687
28	2.3	2.0	421	9	AV788245
29	2.3	2.0	421	9	AV788317
30	2.3	2.0	424	9	AV796683
31	2.3	2.0	424	9	AV820962
32	2.3	2.0	425	9	AV796827
33	2.3	2.0	430	9	AB810526
34	2.3	2.0	431	9	AB810327
35	2.3	2.0	433	9	AV798761
36	2.3	2.0	434	9	AV797571
37	2.3	2.0	438	9	AB802780
38	2.3	2.0	438	9	AA411566
39	2.3	2.0	441	9	AV792423
40	2.3	2.0	442	13	BX096065
41	2.3	2.0	443	9	AB815805
42	2.3	2.0	447	9	AV793780
43	2.3	2.0	455	9	AV794649
44	2.3	2.0	483	9	AV536021
45	2.3	2.0	505	9	AI479358

ALIGNMENTS

RESULT 1
LOCUS CB221243
DEFINITION IABO32A10 Bos taurus Abomasum #1 library Bos taurus CDNA, mRNA
sequence.
ACCESSION CB221243
VERSION CB221243.1 GI:28291757
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
REFERENCE Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Sensen, C.W., Gordon, P.M.K. and Moore, S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0165
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
Insert Length: 307 Std Error: 0.00
POLYA=No.

```

FEATURES
  source      Location/Qualifiers
              1..307
                /organism="Bos taurus"
                /mol_type="mRNA"
                /db_xref="taxon:9913"
                /tissue_type="Smooth muscle"
                /cell_type="Epithelial"
                /dev_stage="Young adult"
                /lab_host="XLI-BlueRF"-strain"
                /clone_1lb="Bos taurus Abomasum #1 library"
                /note="Organ: Abomasum; Vector: Uni-2ZAPXR; Site_1: EcoR
              I; Site_2: Xho I"
              62 a      91 c      93 g      61 t
BASE COUNT
ORIGIN

Query Match      4.2%; Score 49; DB 14; Length 307;
Best Local Similarity 100.0%; Pred. No. 4.9e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      656 TGTTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 704
        |||
        65 TGTTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 113

RESULT 2
LOCUS      BG938086      383 bp      mRNA      linear      EST 11-JUN-2001
DEFINITION      LAB011A08 Bovine Abomasum cDNA library Bos taurus cDNA 5', mRNA
ACCESSION      BG938086
VERSION      BG938086.1      GI:14337458
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
              Bovidae; Bovinae; Bos.
              1 (bases 1 to 383)
              Email: stephen.moore@ualberta.ca
              Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
              CDNA's from bovine abomasum tissue
              Unpublished
              Contact: Dr. Stephen Moore
              Beef Genomics Laboratory
              Dept of AFNS, University of Alberta
              410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
              Tel: 780 492 0169
              Fax: 780 492 4265
              Email: stephen.moore@ualberta.ca
              The sequence best matches gb:BP019786 (Bos primigenius prothymosin
              mRNA, complete cds) in main database at high score of 735.0 and
              E-value of 0.0
              PCR Primers
              FORWARD: M13 Forward
              BACKWARD: M13 Reverse
              Seq primer: T3 primer
              High quality sequence stop: 383
              POLYA=No.
              Location/Qualifiers
                1..383
                  /organism="Bos taurus"
                  /mol_type="mRNA"
                  /db_xref="taxon:9913"
                  /sex="Two males and one female mixed"
                  /tissue_type="Gastrointestinal tissue (GIT)"
                  /cell_type="Epithelial"
                  /dev_stage="Young adult"
                  /lab_host="XLI-BlueRF"-strain"
                  /clone_1lb="Bovine Abomasum cDNA library"
                  /note="Organ: Abomasum; Vector: Uni-2ZAPXR; Site_1: EcoR
                I; Site_2: Xho I"
                83 a      111 c      113 g      76 t
BASE COUNT
ORIGIN

```

```

Query Match      4.2%; Score 49; DB 12; Length 383;
Best Local Similarity 100.0%; Pred. No. 5.2e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      656 TGTTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 704
        |||
        58 TGTTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 106

RESULT 3
LOCUS      BG937697      430 bp      mRNA      linear      EST 11-JUN-2001
DEFINITION      LAB005E01 Bovine Abomasum cDNA library Bos taurus cDNA 5', mRNA
ACCESSION      BG937697
VERSION      BG937697.1      GI:14337069
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
              Bovidae; Bovinae; Bos.
              1 (bases 1 to 430)
              Email: stephen.moore@ualberta.ca
              Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
              CDNA's from bovine abomasum tissue
              Unpublished
              Contact: Dr. Stephen Moore
              Beef Genomics Laboratory
              Dept of AFNS, University of Alberta
              410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
              Tel: 780 492 0169
              Fax: 780 492 4265
              Email: stephen.moore@ualberta.ca
              The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)
              mrnal in main database at high score of 844.0 and E-value of 0.0
              PCR Primers
              FORWARD: M13 Forward
              BACKWARD: M13 Reverse
              Seq primer: T3 primer
              High quality sequence stop: 430
              POLYA=No.
              Location/Qualifiers
                1..430
                  /organism="Bos taurus"
                  /mol_type="mRNA"
                  /db_xref="taxon:9913"
                  /sex="Two males and one female mixed"
                  /tissue_type="Gastrointestinal tissue (GIT)"
                  /cell_type="Epithelial"
                  /dev_stage="Young adult"
                  /lab_host="XLI-BlueRF"-strain"
                  /clone_1lb="Bovine Abomasum cDNA library"
                  /note="Organ: Abomasum; Vector: Uni-2ZAPXR; Site_1: EcoR
                I; Site_2: Xho I"
                97 a      130 c      119 g      84 t
BASE COUNT
ORIGIN

Query Match      4.2%; Score 49; DB 12; Length 430;
Best Local Similarity 100.0%; Pred. No. 5.4e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      656 TGTTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 704
        |||
        52 TGTTCTCGGTTTACATGACAGGAATGCCAGAGACATGCTCAGCCT 100

RESULT 4
LOCUS      BG938320      472 bp      mRNA      linear      EST 11-JUN-2001
DEFINITION      LAB015E12 Bovine Abomasum cDNA library Bos taurus cDNA 5', mRNA
ACCESSION      BG938320
VERSION      BG938320.1      GI:14337692

```

KEYWORDS
EST.
Bos taurus (cow)
SOURCE
Bos taurus
Bos taurus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 472)
Moore, S.S., Hansen, C., Li, C., Fu, A., Meng, Y. and Li, G.
CDNA's from bovine abomasum tissue
JOURNAL
Unpublished
COMMENT
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (yemini)
mrna) in main database at high score of 928.0 and E-value of 0.0
PCR primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 472
POLYA=No.

FEATURES
source
1..472
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF"-strain"
/clone_lib="Bovine Abomasum cDNA library"
/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: Ecor
I; Site_2: Xho I"

BASE COUNT
104 a 143 c 131 g 94 t

ORIGIN
Query Match 4.2%; Score 49; DB 12; Length 472;
Best Local Similarity 100.0%; Pred. No. 5.6e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
656 TGTTCGCGTTTACATGACAGCAATGGCCAGAGACATGCTCAGCCT 704
|||||
65 TGTTCGCGTTTACATGACAGCAATGGCCAGAGACATGCTCAGCCT 113
|||||

Db
65 TGTTCGCGTTTACATGACAGCAATGGCCAGAGACATGCTCAGCCT 113
|||||

RESULT 5
BG937723 399 bp mRNA linear EST 11-JUN-2001
LOCUS
1ABO05D06 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA
sequence.
ACCESSION
BG937723
VERSION
BG937723.1 GI:14337095
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
Bos taurus
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 399)
Moore, S.S., Hansen, C., Li, C., Fu, A., Meng, Y. and Li, G.
CDNA's from bovine abomasum tissue
JOURNAL
Unpublished
COMMENT
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265

Email: stephen.moore@ualberta.ca
The sequence best matches gb:BP019786 (Bos primigenius prochymosin
mRNA, complete cds) in main database at high score of 767.0 and
E-value of 0.0
PCR primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 399
POLYA=No.

FEATURES
source
1..399
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="Two males and one female mixed"
/tissue_type="Gastrointestinal tissue (GIT)"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF"-strain"
/clone_lib="Bovine Abomasum cDNA library"
/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: Ecor
I; Site_2: Xho I"

BASE COUNT
89 a 121 c 112 g 77 t

ORIGIN
Query Match 3.7%; Score 43; DB 12; Length 399;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
662 CGGTTTACATGACAGCAATGGCCAGAGACATGCTCAGCCT 704
|||||
1 CGGTTTACATGACAGCAATGGCCAGAGACATGCTCAGCCT 43
|||||

Db
1 CGGTTTACATGACAGCAATGGCCAGAGACATGCTCAGCCT 43
|||||

RESULT 6
CB220875 410 bp mRNA linear EST 10-FEB-2003
LOCUS
1ABO27D10 Bos taurus Abomasum #1 library Bos taurus cDNA, mRNA
sequence.
ACCESSION
CB220875
VERSION
CB220875.1 GI:28291389
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
Bos taurus
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 410)
Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Senses, C.W., Gordon
P.M.K. and Moore, S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
JOURNAL
Unpublished
COMMENT
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
Insert length: 410 Std Error: 0.00
POLYA=No.

FEATURES
source
1..410
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Smooth muscle"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF"-strain"
/clone_lib="Bos taurus Abomasum #1 library"
/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: Ecor
I; Site_2: Xho I"

BASE COUNT 90 a 123 c 120 g 77 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 663 GGTTCATGACGAGCATGGCCAGAGACATGCTCAGCCT 704
DB 36 GGTTCATGACGAGCATGGCCAGAGACATGCTCAGCCT 77

RESULT 7
CB220520 297 bp mRNA linear EST 10-FEB-2003
LOCUS 1Ab0231a12 Bos taurus Abomasum #1 library Bos taurus cDNA, mRNA
DEFINITION
sequence.
ACCESSION CB220520
KEYWORDS CB220520.1 GI:28291034
SOURCE EST.
ORGANISM Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 297)
Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Sensen, C.W., Gordon
, P.M.K. and Moore, S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
Insert Length: 297 Std Error: 0.00
POLYA=No.

FEATURES
source Location/Qualifiers
1..297
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Smooth muscle"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMRP-strain"
/clone_1lb="Bos taurus Abomasum #1 library"
/note="Organ: Abomasum; Vector: Uni-2ZAPXR; site_1: Ecor
I; Site 2: Xho I"

BASE COUNT 67 a 91 c 81 g 58 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 778 TTCACGTGAGCGTTCACATCAGCGGTGTTGT 815
DB 16 TTCACGTGAGCGTTCACATCAGCGGTGTTGT 53

RESULT 8
CB220074 443 bp mRNA linear EST 10-FEB-2003
LOCUS 1Ab017B12 Bos taurus Abomasum #1 library Bos taurus cDNA, mRNA
DEFINITION
sequence.
ACCESSION CB220074
KEYWORDS CB220074.1 GI:28290568
SOURCE EST.
ORGANISM Bos taurus (cow)

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 443)
Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Sensen, C.W., Gordon
, P.M.K. and Moore, S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished
Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: stephen.moore@ualberta.ca
Insert Length: 443 Std Error: 0.00
POLYA=Yes.

FEATURES
source Location/Qualifiers
1..443
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="Smooth muscle"
/cell_type="Epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMRP-strain"
/clone_1lb="Bos taurus Abomasum #1 library"
/note="Organ: Abomasum; Vector: Uni-2ZAPXR; site_1: Ecor
I; Site 2: Xho I"

BASE COUNT 125 a 132 c 96 g 90 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 907 ATTGAGCCACACAGACGTAACGTTGACAT 944
DB 37 ATTGAGCCACACAGACGTAACGTTGACAT 74

RESULT 9
B2911661 204 bp DNA linear GSS 12-JUN-2003
LOCUS CH240.109D12.TV CHOR1-240 Bos taurus genomic clone CH240_109D12,
DEFINITION genomic survey sequence.
ACCESSION B2911661
KEYWORDS B2911661.1 GI:31637047
SOURCE GSS.
ORGANISM Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 204)
Larkin, D.M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P.,
Bachman, S., Green, S., Campos, E.J., Benson, L.D., Edwards, J., Liu, L.,
Mowack, J.E., de Jong, P.J. and Lewin, H.A.
A Cattle-Human Comparative Map Built with Cattle BAC-ends and Human
Genome Sequence
Unpublished
Other GSSs: CH240.109D12.TV
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Clones are derived from the bovine BAC library CHOR1-240
(http://www.cho1.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/ordering/information.htm>). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBMC) by the University of Illinois at Urbana-Champaign, USA with funds provided by grant No. AG202-34480-11828 from USDA-CSRES and AG99-35205-8534 from USDA/NRI (Livestock Genome Sequencing Initiative)
 Plate: 109 row: D column: 12
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Source

Location/Qualifiers
 1. .204
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="bred: Hereford"
 /db_xref="taxon:9913"
 /clone="CH240_109D12"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: pTRABAC1.3; Site 1: Mbol; Site 2: Mbol; Hereford bull Li Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"
 BASE COUNT 42 a 60 c 54 g 43 t
 ORIGIN

Query Match 2.7%; Score 32; DB 29; Length 204;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 CAGTATGGCATCAGCAGCACTACCGCGCTT 200
 |||||
 22 CAGTATGGCATCAGCAGCACTACCGCGCTT 53

RESULT 10
 AV532492/c 325 bp mRNA linear EST 01-SEP-2000
 LOCUS AV532492 Arabidopsis thaliana flower buds Columbia Arabidopsis
 DEFINITION thaliana cDNA clone FB043b03f 3', mRNA sequence.
 ACCESSION AV532492
 VERSION AV532492.1 GI:8692775
 KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 325)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)

JOURNAL 20363093
 MEDLINE 10907847
 PUBMED
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES
 source Location/Qualifiers
 1. .325
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="FB043b03f"
 /tissue_type="flower buds"
 /clone_lib="Arabidopsis thaliana flower buds Columbia"
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI"

BASE COUNT 106 a 92 c 59 g 68 t

ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 325;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGAGATGTGTTTCAT 1109
 |||||
 DB 209 TGGATCTTGGAGATGTGTTTCAT 187

RESULT 11

T20903 339 bp mRNA linear EST 07-JAN-1998
 LOCUS T20903
 DEFINITION 2911 lambda-PRL2 Arabidopsis thaliana cDNA clone 89P9T7, mRNA sequence.

ACCESSION T20903
 VERSION T20903.1 GI:2756823
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE Newman, T., deBruin, J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M., Retzel, E., and Somerville, C.
 Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 Plant Physiol. 106, 1241-1255 (1994)

JOURNAL 95148729
 MEDLINE 7846151
 PUBMED
 COMMENT On Jan 7, 1998 this sequence version replaced gi:502344.
 Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313cne@lhm.cl.msu.edu
 Seg primer: T7.

FEATURES
 source Location/Qualifiers
 1. .339
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="varolumbia"
 /db_xref="taxon:3702"
 /clone="89P9T7"
 /note="Vector: lambda-PRL2"
 /note="Vector: lambda 2ip-lox; Site 1: Sal; Site 2: Not; lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA."

BASE COUNT 67 a 61 c 92 g 106 t 13 others
 ORIGIN

Query Match 2.0%; Score 23; DB 14; Length 339;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGAGATGTGTTTCAT 1109
 |||||
 DB 82 TGGATCTTGGAGATGTGTTTCAT 104

RESULT 12
 AVS38262/c 380 bp mRNA linear EST 06-SEP-2000
 LOCUS
 DEFINITION Arabidopsis thaliana roots Columbia Arabidopsis thaliana
 ACCESSION
 AVS38262
 VERSION
 AVS38262.1 GI:8698545
 KEYWORDS
 EST.
 SOURCE
 Arabidopsis thaliana (thale cress)
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 380)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL
 DNA Res. 7, 175-180 (2000)
 MEDLINE
 20365093
 PUBMED
 10907847
 COMMENT
 Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 Location/Qualifiers
 1..380
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="R211a03f"
 /tissue_type="roots"
 /clone_1ib="Arabidopsis thaliana roots Columbia"
 /note="Vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT
 130 a 96 c 65 g 89 t
 ORIGIN
 Query Match 2.0%; Score 23; DB 9; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1087 TGGATCTTGGAGATGTTTCAT 1109
 |||||||
 |||||||
 Db 332 TGGATCTTGGAGATGTTTCAT 310
 RESULT 13
 AA969042/c 391 bp mRNA linear EST 07-JUL-1998
 LOCUS
 DEFINITION Arabidopsis thaliana roots Columbia Arabidopsis thaliana
 ACCESSION
 AA969042
 VERSION
 AA969042.1 GI:3144222
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 391)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgaf.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-f@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 619 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham

FEATURES
 source
 High quality sequence stop: 378.
 Location/Qualifiers
 1..391
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1579597"
 /lab_host="DH10B"
 /clone_1ib="Soares_NFL_T_GBC_S1"
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CCAP GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 725096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 726408-728711, and 725096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT
 82 a 101 c 115 g 93 t
 ORIGIN
 Query Match 2.0%; Score 23; DB 9; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1015 CCCTCCGCTATACGACGAGA 1037
 |||||||
 |||||||
 Db 256 CCTCCGCTATACGACGAGA 234
 RESULT 14
 AV801400/c 396 bp mRNA linear EST 29-MAR-2002
 LOCUS
 DEFINITION Arabidopsis thaliana cDNA clone RAFL09-27-P07 3',
 mRNA sequence.
 ACCESSION
 AV801400
 VERSION
 AV801400.1 GI:19835385
 KEYWORDS
 EST.
 SOURCE
 Arabidopsis thaliana (thale cress)
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 396)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saito, M., Nakajima, M.,
 Oono, Y., Sakurai, T., Carninci, P., Kawai, T., Itoh, M., Ishii, Y.,
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
 and Shinozaki, K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Unpublished
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
 and XhoI was ligated to modified lambda PUC-1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified pBluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
 details.
 Location/Qualifiers
 1..396
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL09-27-P07"

/dev stage="plants at various developmental stages from
germination to mature seeds"
/lab host="DH10B"
/clone lib="RAFL9"
/note="Site 1: BamHI; Site 2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT 133 a 101 c 70 g 92 t
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGGAGATGTTTCAT 1109
|||||
DB 330 TGGATCTTGGGAGATGTTTCAT 308

RESULT 15
AV801901/c 398 bp mRNA linear EST 29-MAR-2002
LOCUS AV801901 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-29-P20 3',
DEFINITION mRNA sequence.
ACCESSION AV801901 GI:19835886
VERSION AV801901.1 GI:19835886
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; euroside II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 398)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saitou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekier@cc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda PhiC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source 1..398
Location/Qualifiers

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-29-P20"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab host="DH10B"
/clone lib="RAFL9"
/note="Site 1: BamHI; Site 2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"

BASE COUNT 133 a 103 c 71 g 91 t
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 398;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 TGGATCTTGGGAGATGTTTCAT 1109
|||||
DB 332 TGGATCTTGGGAGATGTTTCAT 310

Search completed: February 5, 2004, 08:12:44
Job time : 2894 secs

THIS PAGE BLANK (USPTO)